Numerical production of vowels and diphthongs using finite element methods

by
Marc Arnela Coll

A thesis submitted for the degree of Doctor of Philosophy

Supervised by Dr. Oriol Guasch i Fortuny

La Salle, Universitat Ramon Llull,
Barcelona, 2014.

Marc Arnela Coll

Numerical production of vowels and diphthongs using finite element methods

Programa de doctorat en
"Les Tecnologies de la Informació i les Comunicacions i la seva gestió",
La Salle, Universitat Ramon Llull.

Supervised by Dr. Oriol Guasch i Fortuny

GTM Grup de recerca en Tecnologies Mèdia,
La Salle, Universitat Ramon Llull.

Marc Arnela, "Numerical production of vowels and diphthongs using finite element methods," Ph.D. dissertation, GTM Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull, Barcelona, Catalonia.

Abstract

In recent times, three-dimensional computer simulations have become quite popular to model the acoustics of the vocal tract. They can overcome some of the classical limitations of one-dimensional techniques and include all intricate details of the vocal tract geometry, greatly improving the quality of the generated sounds. The finite element method (FEM) is probably the most appropriate approach for solving the underlying equations of voice.

In the first part of this thesis the problem of vowel production is addressed. FEM is used to solve the wave equation for the acoustic pressure combined with a Perfectly Matched Layer (PML) to account for free-field radiation conditions. This allows the simulation of acoustic waves propagating through the vocal tract and emanating from the mouth aperture. The proposed FEM approach is then validated against experiments performed with simplified vocal tract replicas. Following is an adaptation of a two-microphone transfer function method to compute vocal tract impedances, which becomes later used to analyze the radiation effects of human head simplifications on vowels. Finally, it is proposed to perform two-dimensional (2D) simulations based on a tuning process that allows 2D vocal tracts to mimic the acoustics of 3D vocal tracts, to a large extent. This results in a very good balance between computational cost and voice quality. In the second part of the thesis diphthong sounds are generated. A stabilized finite element scheme for the mixed wave equation in an Arbitrary Lagrangian-Eulerian (ALE) framework is derived for that purpose. Diphthongs are produced using both, 3D moving vocal tracts as well as 2D tuned vocal tracts.

The proposed approach for vowels and diphthongs allows not only to visualize waves propagating within the vocal tract but also to listen to the corresponding generated sounds.

Acknowledgements

Voldria dedicar les primeres paraules d'agraïment al meu director de tesi, el Dr. Oriol Guasch i Fortuny. Gràcies per la teva infinita paciència i dedicació constant durant tots aquests anys. Sempre has tingut un moment per resoldre els meus dubtes. En segon lloc, al Dr. Francesc Alías Pujol, que tot i que finalment no figura com a codirector d'aquesta tesi, el treball realitzat en la present no hagués estat possible.

Miro enrere i ja fa molts anys que em vaig iniciar en el camp de l'acústica. Tot va començar en el encara departament d'acústica de la Salle, on vaig entrar per fer un any d'intensificació i en van acabar seguint tres més com a becari de docència, donant suport a l'assignatura de mesures acústiques. Voldria agrair doncs als meus companys de departament, la Ivana, el Pere, el Sergi, el Matthias, l'Umut i el Ricardo, per ajudar-me i fer-me atractiu aquest camp i confiar en mi en els meus inicis. Voldria tenir unes especials paraules d'agraïment al Lluís, amb qui vaig compartir despatx i diverses batalles durant tots aquests anys, i també a l'Ester i al Carlos, els quals em van succeir en les meves tasques. També a aquells companys d'universitat que em van enriquir com a persona durant aquesta etapa. Pepu, Ana, Víctor, Cris i Ester, moltes gràcies. Espero que continuem fent els nostres sopars al Fosters molts més anys.

Poc més tard ja vaig començar el doctorat en la mateixa institució, una aventura que duraria gairebé cinc anys i mig i en la que abandonava el món experimental en el que em sentia tant còmode per endinsar-me en el desconegut camp de l'acústica computacional. És molta la gent que de manera directa o indirecta hi ha contribuït. Voldria agrair en primer lloc a tots els membres del selecte club "Esmorzar pagant", amics a la feina i companys de viatge. Àngels, Raúl, Marc, Ramon, Xavi V., Xavi S., Xuti, Adso i Alan, ha estat tot molt més fàcil gràcies a vosaltres. Aquest llarg procés s'hauria tornat més difícil de suportar si no fos pels nostres esmorzars i dinars plegats, incloent per suposat les nostres tertúlies i discussions. També voldria agrair als amics de sempre, els de l'escola, amb els quals he crescut i m'he format com a persona, i també a aquells que m'he anat trobant pel camí. Samir, Joan, Dani, Iñaki, Josep, Mario, Marcos, Olga, Gemma, Sofi,

Vane i Cris, gràcies per donar-me forces tots aquests anys. Heu estat presents tant en els bons com en els mals moments. La realització del doctorat tampoc hagués estat possible sense una de les meves principals passions, la música. A tots els membres del cor de cambra "En clau vocal", Elena, Núria, Chus, Carmina, Marta, Lupe, Jose, Andrés, Martin, gràcies per tots els bons moments compartits. Espero que continuem fent bona música molts anys més. També agrair a les meves professores de piano Mireia i Iwona, les quals em van obrir la porta a aquest apassionant món i em van ensenyar a estimar-lo.

Les últimes paraules en català les voldria tenir per la meva família. El meu pare, la meva mare, el meu germà Lluís, la meva cunyada Noe i els meus nebots Aina i Arnau que tantes alegries m'han donat. El vostre suport incondicional durant tots aquests anys no té preu. Gràcies de tot cor.

I would like to also acknowledge Prof. Olov Engwall and Prof. Sten Ternström for hosting me during the seven months that my research stay lasted at the Department of Speech, Music and Hearing (TMH) at the Royal Institute of Technology (KTH), Stockholm. Jag har haft en fantastisk tid där i TMH. Tack så mycket Sten for introducing me to Danderyds Vokalensemble, with whom I spent great musical moments, and for giving me access to the grand piano in the basement. Special thanks to Saeed Dabbaghchian, my office mate, it was so easy to work with you, and also to Andreas Selamtzis, for useful discussions about voice production research and for showing me folkmusikhuset in Skeppis, "the folk music place". Thanks Catha for all those nice moments (and little adventures), and also to the rest of the people in the department, who were very kind to me. During this research stay, the MRI-based geometries used in Chapter 2 were constructed together with Saeed Dabbaghchian and Prof. Olov Engwall. I really appreciate all this joint work. Tack igen!

I am very grateful to Prof. Ramon Codina, from the International Center for Numerical Methods in Engineering (CIMNE) in Barcelona, for providing us with an initial FEM code for the scalar convection-difusion-reaction equation from which the code for vowel production was developed (Chapter 2), and also to Héctor Espinoza and Prof. Ramon Codina for sharing a FEM code for the wave equation in mixed form, which was adapted and extended to simulate diphthong sounds (Chapter 8). I would like to also acknowledge Rémi Blandin, Prof. Xavier Pelorson and Prof. Annemie Van Hirtum, from GIPSA-lab in Grenoble, for supplying us with the experimental data used to validate in Chapter 3 the FEM approach for vowel generation.

I would like to acknowledge the Generalitat de Catalunya (SUR/ECO) for the pre-doctoral FI grants No. 2012FI_B 00727, No. 2013FI_B1 00165 and No. 2014FI_B2 00142, and the exchange BE Grant No. $2012 - \text{BE}100242$. This work has been partially funded by EU-FET grant EUNISON 308874.

Table of Contents

Part I

Overview

Chapter 1

Introduction

1.1 Motivation and scope

In the past decades several strategies have been followed to simulate human voice. For historical reasons, various simplifications have been attempted in order to generate realistic voice signals. For instance, the pre-recorded speech segment concatenation or parametric modelling (e.g., Hidden Markov Models or Harmonic plus Noise Models) that current speech synthesizers incorporate, or the several tricks that classical articulatory models make when approximating the intricate three-dimensional (3D) physics of voice to more simplified one-dimensional strategies. However, the amazingly growing capacity of computers combined with extensive research on numerical mathematics and medical imaging, has opened the door to go one step beyond and simulate the whole 3D mechanism of voice production. This means to simulate from the vocal folds vibration and the glottal flow they generate, to its filtering by the vocal tract and final voice emission to free space. Several physical phenomena have to be considered to do so, which include the interaction between the mechanical, aerodynamic and acoustic fields, the collision of the vocal folds, the generation of turbulent airflows and the propagation of acoustic waves in a dynamic vocal tract.

It is at the aim of this work to contribute to the construction of this three-dimensional voice simulation engine. In particular, to address the problem of vowel and diphthong production by numerically solving the involved equations using Finite Element Methods. Special attention is paid to the vocal tract acoustic modelling, which constitutes the core of this thesis.

The applications of such a voice engine are wide in the mid-long term, in addition to the basic knowledge that it may provide to the better understanding on how voice works (and fails). These range from synthesizing natural and personalized speech not depending on pre-recorded speech corpora, to improved medical procedures (e.g., simulating the acoustic effects of a surgery), pedagogy and education of voice (e.g., improvement of second language learning by visualizing the voice organs and the generated acoustic waves while listening the produced sound), media technologies (e.g., a human avatar with a real synchrony between voice and image) and new forms of cultural expression.

1.2 A brief review of articulatory voice models

Any articulatory model for voice production has to address three issues: having an appropriate geometrical model of the voice organ, having an acoustic source model for the pulses generated at the vocal folds and having an adequate model for the vocal tract acoustics. Let us briefly review their current state of the art with special emphasis on the latter, which constitutes the core of the present thesis.

1.2.1 Block diagram

An articulatory voice model can be divided into the following parts conforming to the block diagram in Fig. 1.1:

- **Glottal source model:** the main goal of this block is to mimic the behavior of the phonatory organs (vocal folds) to generate the input inflow into the vocal tract. For voiced sounds (e.g. vowels and diphthongs), this airflow corresponds to a train of pseudoperiodic pulses known as glottal pulses.

- **Vocal tract geometry model:** this model generates the vocal tract geometry corresponding to the sound produced at each instant of time. For vowels a static configuration of the vocal tract can be assumed while for diphthongs dynamic vocal tracts will be required.

- **Vocal tract acoustic model:** using as input the vocal tract geometries and the train of glottal pulses generated by the vocal folds, the time evolution of the acoustic pressure is computed within the vocal tract and in the near free-field. Tracking then the acoustic pressure close to the mouth aperture and transforming it to an audio file we get the generated sound. For instance, if the geometry model provides a

set of vocal tracts transiting from vowel /a/ to vowel /i/, the produced sound will correspond to the diphthong /ai/.

- **Articulatory control:** this is the "front-end" of the articulatory model and it is the responsible to control all the above models for generating a given sound. Simple configurations will be given to this model. However, in the future, it could be replaced for example with a more sophisticated text-to-phonetics-to-neuromotor "front-end".

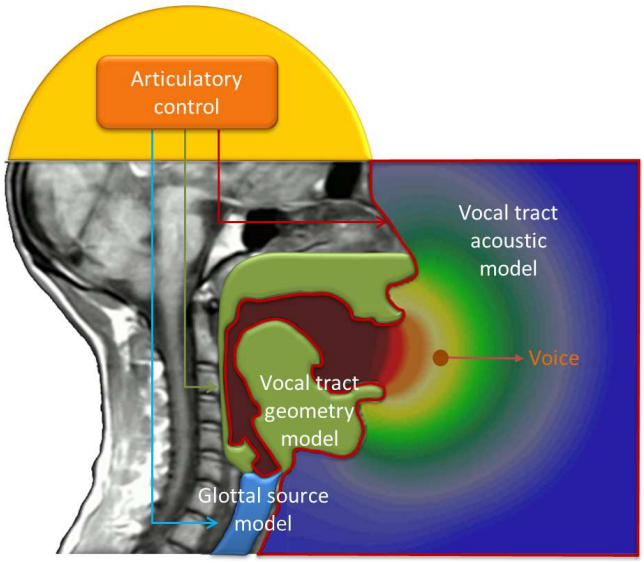
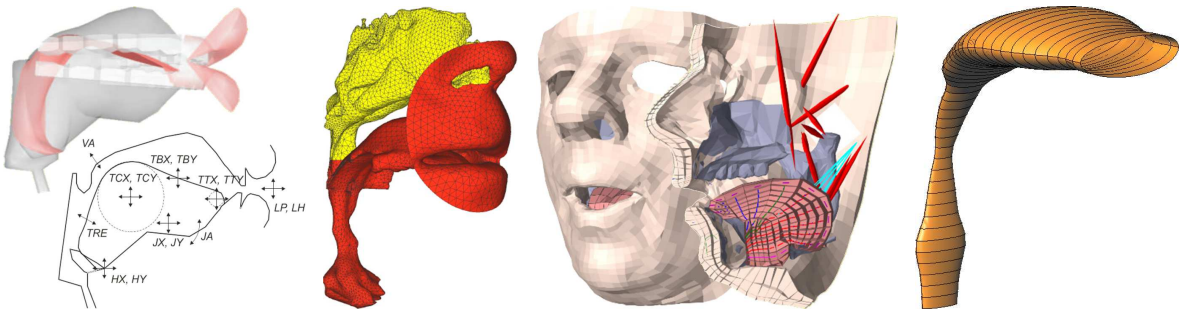


Fig. 1.1: Block diagram of the proposed time domain FEM articulatory voice model.

As previously commented, this work is focused on the vocal tract acoustic modelling, and specifically on the use of finite element methods to solve the equations that describe the production of vowels and diphthongs. In what follows, we will briefly describe the main strategies used to model each block in Fig. 1.1 and justify the options selected in this thesis.

1.2.2 Vocal tract geometry model

The first thing we need is a geometry that resembles the human vocal tract, in order to lately compute its acoustic behavior. That is, we need to know the inner boundary of the vocal tract cavity (which implies knowing the shapes and positions of all vocal tract articulators such as lips, jaw, tongue, palate, nasal cavity, etc.) when a given sound is produced. If the sound is not steady, then the geometry will evolve with time (e.g., when pronouncing a diphthong or a syllable) and, hence, this evolution also has to be modeled.



(a) (b) (c) (d)

Fig. 1.2: Vocal tract geometry models. (a) Geometrical model by Birkholz and Kröger (2006), (b) statistical model by Matsuzaki *et al.* (2008), biomechanical model generated using Artisynth environment by Stavness *et al.* (2011), and (d) 3D vocal tract airway of vowel /a/ reconstructed by the author from area functions in Story (2008). The latter is the option used through this work for the vocal tract geometry model.

Depending on how the vocal tract geometry is approximated, the corresponding models can be classified as being geometrical (see Fig. 1.2a), statistical (see Fig. 1.2b) or biomechanical (see Fig. 1.2c) (Kröger and Birkholz, 2009):

- Geometrical models (e.g., Mermelstein, 1973; Birkholz *et al.*, 2006) rely on simulating the complex vocal tract airflow linking simple geometric elements (i.e., circumferences, arcs, squares, etc.). The various parameters of these elements (i.e. radius, dimensions, etc.) can be modified to simulate the articulatory movements of the vocal tract. Geometric models are the most flexible ones and can be adapted to mimic any speaker's vocal tract (different age and sex, Birkholz and Kröger, 2006).

- Statistical models (e.g., Stark *et al.*, 1999; Serrurier and Badin, 2008; Palo *et al.*, 2012; Aalto *et al.*, 2014) obtain the vocal tract geometry from huge databases measured using different techniques such as MRI (Magnetic Resonance Imaging), CT (Computed Tomography), X-ray or/and EMA (a review of these techniques applied to voice production can be found in Ridouane, 2006). Although very precise and realistic, statistical models just reproduce the characteristics of a single speaker and become much less flexible than geometrical models. Their advantage is that they deal with a relatively small set of uncorrelated parameters.

- Biomechanical models (e.g., Dang and Honda, 2004; Fels *et al.*, 2009; Stavness *et al.*, 2011) make use of numerical methods (typically FEM) to simulate the dynamics of the vocal tract. Physiological knowledge on the relation between muscle activation and articulatory movements for voice synthesis is therefore required. These models usually involve a large number of parameters and are difficult to control.

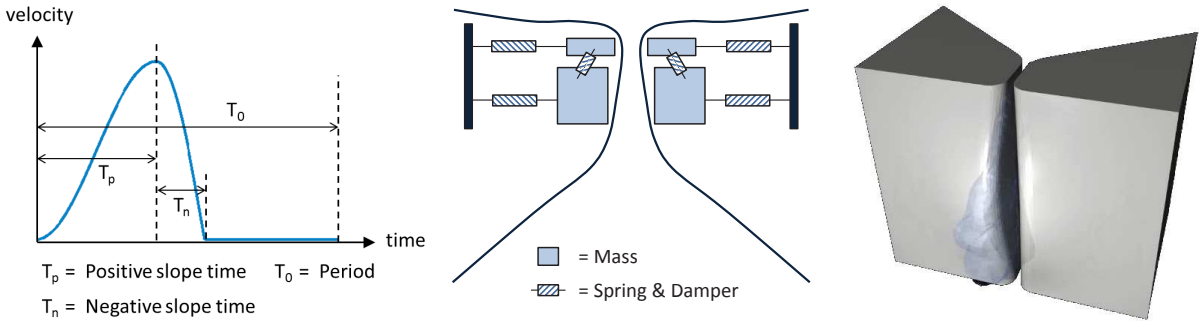
The above described models generate three-dimensional geometries. However, some vocal tract acoustic models deal with faster one-dimensional simulations. To do so, a process is performed consisting in computing the area of each cross-section perpendicular to the midline of the vocal tract airway, at each time instant (see e.g., Birkholz and Jackèl, 2003). As a result, a time-varying area function describing the vocal tract geometry is obtained. Several data sets for the area functions of different sounds can be found in the literature (see e.g., Story *et al.*, 1996; Story, 2008, where area functions are obtained from MRI measurements).

For the present work, area functions available in the literature (such as Story, 2008) will be used for simplicity, from which 3D vocal tracts can be reconstructed (see Fig. 1.2d). Interpolation strategies will be performed to generate dynamic vocal tract geometries from static ones, so as to simulate diphthongs. The election of this simplified geometries will be latter discussed in Section 2.5.1, where detailed vocal tracts obtained from MRI will be compared to their simplified counterparts.

1.2.3 Glottal source model

Once defined the vocal tract geometry and prior to the simulation of its acoustics response, we need to know how sound is generated in it. This is the main goal of the glottal source models. As previously commented, for voiced sounds (e.g. vowels and diphthongs), a train of pseudoperiodic pulses known as glottal pulses is generated by the vocal folds, which act on the steady airflow coming from the lower respiratory tract (trachea, lungs, etc.). Basically, glottal source models can be divided into waveform models (see Fig. 1.3a), self-oscillating models (see Fig. 1.3b) and computational models (see Fig. 1.3c):

- Waveform models approximate the velocity waveform of the airflow generated by the phonatory organs by means of trigonometric functions. The input parameters of these models coincide with some articulatory parameters (e.g. fundamental frequency or pitch, amplitude, etc.). The most celebrated waveform models are that of Rosenberg (Rosenberg, 1971) and that of Liljencrants and Fant (Fant *et al.*, 1985). The latter is also known as LF model.
- In the case of self-oscillating models (see e.g. Story, 2002, for a review), the vocal cords behavior is modeled by means of a mechanical analogy (e.g., coupled mass-spring systems). The system solution gives place to self-sustained oscillations determining the glottal aperture (aperture between the vocal folds, aka glottis) and the glottal waveform (glottal pulses). Self-oscillating models are controlled by biomechanical parameters like the air pressure provided by the lungs, the tension of



(a) (b) (c)

Fig. 1.3: Glottal source models. (a) Waveform model of the Rosenberg type (Rosenberg, 1971) (b) Self-oscillating two-mass model of Ishizaka and Flanagan (1972) (figure adapted from Story, 2002) and (c) Computational FE model of the vocal folds with fluid-structure interaction (figure from Guasch *et al.*, 2013).

the vocal cords, etc. The simplest of these models are the one-mass model (Flanagan and Landgraf, 1968) and the two-mass model (Ishizaka and Flanagan, 1972; Sondhi and Schroeter, 1987; Pelorson *et al.*, 1994; Lous *et al.*, 1998).

- With regard to computational models, they attempt to simulate the fluid mechanics and aeroacoustics of the vocal folds. For instance, in Zhao *et al.* (2002a,b), a direct numerical simulation of the compressible Navier-Stokes equations using a finite difference scheme was carried out. In Tian *et al.* (2014) fluid-structure interaction is addressed in three-dimensional vocal folds by using FEM. However, it has not been until very recently, that a finite element method to solve the coupled equations for the mechanics, fluid dynamics and the acoustics of a two-dimensional glottal system has been presented (Link *et al.*, 2009). Analytical approaches based on the Green's function solution of the corresponding partial differential equations have been also attempted (Howe and McGowan, 2007; McGowan and Howe, 2007). These approaches have the advantage of easy parameter space exploration, but are only suitable for very simplified geometries.

It is worthwhile noting that although LF and two-mass models were developed quite a long time ago, they are still very popular as they combine simplicity with very acceptable results (see e.g., Doel and Ascher, 2008). Our choice for the present work will then be that of implementing waveform models such as the Rosenberg or the LF ones, which can be input as boundary conditions for the simulations of vocal tract acoustics. Both models are detailed in Appendix B.

1.2.4 Vocal tract acoustic model

Once having the vocal tract geometry and the glottal source model (vocal tract inflow), we can focus on simulating the acoustics of the vocal tract, which will finally yield the generated voice. Vocal tract acoustic models can be divided in those that rely on one-dimensional strategies and those that follow a three-dimensional approach.

One-dimensional models

For historical reasons and thanks to their simplicity, one-dimensional models (aka tube models) have become widely known and used. These models approximate the vocal tract geometry as a finite set of concatenative tubes, each one having constant cross-section (i.e., vocal tract area functions, e.g. Story *et al.*, 1996). Then, assuming plane wave propagation within the vocal tract, which holds up to about 5 kHz, the acoustic pressure at the exit (voice) is computed. They can be subdivided into the ABCD matrix based models (see Fig. 1.4a), the Digital Waveguides Models (DWM, see Fig. 1.4b), circuit analogy models (see Fig. 1.4c) and computational models:

- Tube models of the ABCD matrix type (e.g., Sondhi and Schroeter, 1987) compute the acoustic transfer function of the vocal tract as the products of the individual transfer functions for each tube. Then, the voice signal is computed in the frequency domain using the vocal tract transfer function, and the inflow provided by the source model. In contrast to time domain models, ACBD matrix based models

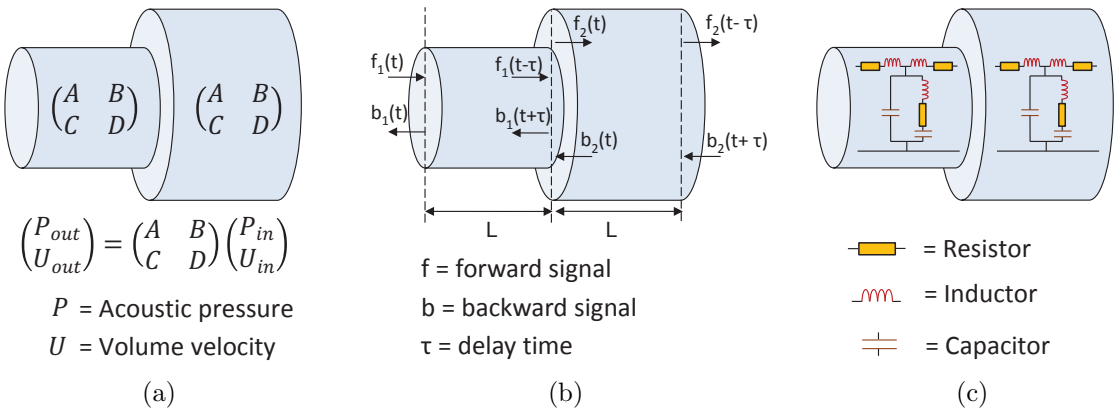


Fig. 1.4: Vocal tract acoustic models based on tubes (1D). (a) ACBD matrix based model, (b) Digital Waveguide Model or KL model and (c) circuit analogy model.

do not directly provide the acoustic pressure or the acoustic velocity within the vocal tract. Consequently, time-frequency transformations are required to obtain their time variations. These models can also be refereed as hybrid time-frequency domain models.

- In Digital Waveguide Models, aka as KL models by Kelly and Lochbaum (1962), the d'Alembertian solution (backward and forward signals) of the one-dimensional acoustic equation is used to describe the acoustic wave propagation within the vocal tract. This allows to approximate the vocal tract as a set of transmission lines. The main problem of these models is that the geometry cannot be changed smoothly (Doel and Ascher, 2008), which is essential for coarticulation processes such as the generation of diphthongs.

- In tube models based on circuit analogies (see e.g., Birkholz *et al.*, 2007; Serrurier and Badin, 2008), the acoustic velocity and pressure waves are respectively interpreted as intensity and voltage signals, and the acoustic properties of each tube are modeled by an electrical circuit representation. These models compute the voice signal in the time domain, but they need to do a lot of approximations in the electric analogy process (Kröger and Birkholz, 2009) (this is the case for instance, of the electrical analogue for the propagation of frequency-dependent waves into free space, which is necessary to account for the propagation of voice emanating from the mouth). In contrast to KL models, circuit models can assume time-varying geometry lengths (Kröger and Birkholz, 2009).

- In 1D computational models the one-dimensional equations that describe the acoustic waves within the vocal tract are solved using numerical methods. Typically finite differences are used for both time and space. For instance, an elaborated version of the Webster equation is solved in Doel and Ascher (2008) to synthesize vowels and diphthongs in real time. The use of these type of approaches is not as extended as the above models. This could be attributed to the fact that computational techniques have been applied to voice production quite recently and maybe that they are more intricate.

Three-dimensional models

Three-dimensional techniques offer wider possibilities than one-dimensional models. Complex vocal tract geometries can be implemented in full detail, plane wave assumption is no longer necessary and some physical phenomena produced during voice production can be taken into account in a more natural way (e.g., aeroacoustics involved in the generation of many sounds). Three-dimensional approaches can be classified in computational models, digital waveguide models and multimodal models:

- Computational models make use of numerical methods to solve the underlaying equations of voice. In the case of vowels and diphthongs, the equation to be solved correspond to the time domain acoustic wave equation either in irreducible form (acoustic pressure) or mixed form (acoustic pressure and particle velocity), or its frequency counterpart, the Helmholtz equation. Computational models can then be subdivided depending on the numerical method that they use and if they work in the frequency or time domain,
  - Time Domain Finite Element Methods (TD-FEM), see e.g. Švancara and Horáček (2006); Švancara *et al.* (2006); Vampola *et al.* (2008a,b, 2011, 2013); Arnela and Guasch (2013); Arnela *et al.* (2013); Arnela and Guasch (2014b),
  - Time Domain Finite Differences (TDFD), see e.g. Takemoto *et al.* (2010, 2013),
  - Frequency Domain Finite Element Methods (FD-FEM), see e.g. Matsuzaki *et al.* (2000); Motoki (2002); Kako and Touda (2006); Hannukainen *et al.* (2007); Zhou *et al.* (2008); Fleischer *et al.* (2014),
  - Frequency Domain Boundary Element Methods (FD-BEM), see e.g. Kagawa *et al.* (1992).

  Working in the time domain seems to be the best solution to describe the voice production mechanism, given that most of the involved physical phenomena also evolve in time (e.g., dynamic vocal tracts, self-oscillation of the vocal folds, etc.). On the other hand, FEM is a more appropriate numerical method than finite differences. It can better discretize complex geometries such as the detailed vocal tracts generated from MRI. In finite differences structured cubic meshes are required, which find difficulties when discretizing complex boundaries (e.g., a curved surface), while in FEM for instance unstructured tetrahedral meshes with arbitrary element size can be used instead. Therefore, the most suitable option among the above numerical method seems to be the Time Domain Finite Element Method.

- The classical theory of Digital Waveguide Models has been recently extended to 3D. Some examples can be found in Speed *et al.* (2013) for simplified vocal tract geometries with circular cross-sections. As for the TDFD, this approach also suffers from a poor spatial discretization of complex geometries since they also require structured cubic meshes.

- In multimodal methods the three-dimensional acoustic field within a succession of constant cross-section waveguides that conform the vocal tract is decomposed in an infinite set of propagation modes. Applying the corresponding continuity equations to each junction, the acoustic waves and therefore the voice signal is analytically derived. This strategy was applied for instance in Motoki *et al.* (2000) to vocal tracts with rectangular shape. More recently, this theory has been extended to vocal tracts witch circular and elliptical shapes and with centric and eccentric junctions (Blandin *et al.*, 2014a, see also Chapter 3 for these geometries). Although these models do not require large computational resources, their application to realistic vocal tracts (with complex shapes) is still a challenge, since their difficulty rely on obtaining the modal basis.

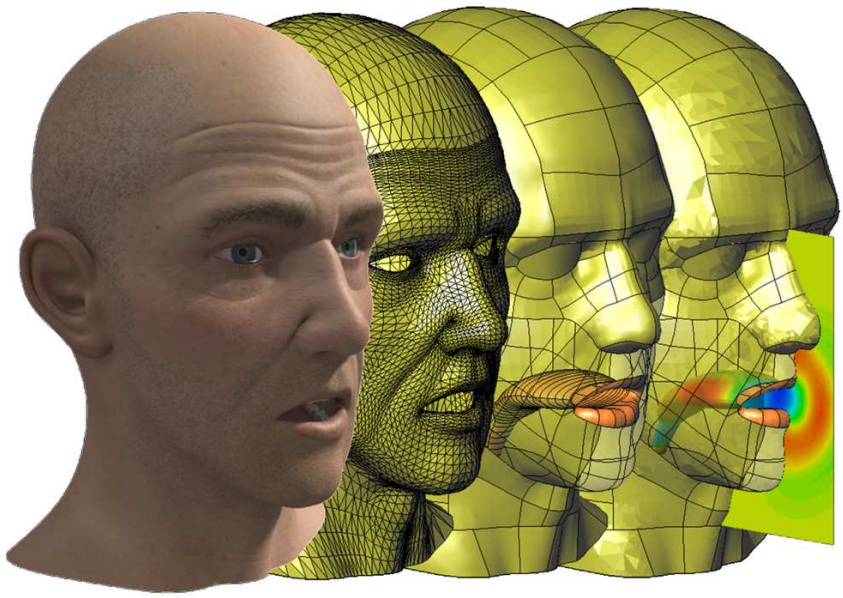


Fig. 1.5: Computational vocal tract acoustic model proposed in this work based on the Time-Domain Finite Element Method (TD-FEM). A realistic human head combined with a simplified vocal tract (see Fig. 1.2d) is used as a vocal tract geometry model (see Chapter 6).

In this work the Finite Element Method will be used to numerically solve the acoustic wave equation in the time domain (TD-FEM) (see Fig. 1.5 for an example). As previously commented, it allow us to better account for the time evolving phenomena of voice and consider intricate geometries of the vocal tract. Moreover, and in contrast to one-dimensional models, three-dimensional acoustical effects can be simulated greatly improving the voice quality, but at the price of increasing the computational times. Some further steps will be specially done with respect to the above cited works when simulating diphthongs, since they mainly deal with static geometries. The production of diphthongs, which implies vowel coarticulation, will require working with moving geometries. Hence, the involved wave equations will have to be expressed in an ALE (Arbitrary Lagrangian-Eulerian) frame of reference (see e.g. Hirt *et al.*, 1974; Hughes *et al.*, 1981; Wall *et al.*, 2006, for more recent advances), which will pose new numerical challenges that will be overcome in this thesis.

On the other hand, some comments should be made concerning numerical issues related to FEM. It is well-known that the standard Galerkin approach to solve partial differential equations may suffer from several numerical instabilities (e.g., the Galerkin weak form for the Helmholtz equation suffers from the so-called pollution error as it becomes non-positive definite for large wavenumbers). In the past decades, several numerical strategies have been derived to circumvent these numerical instabilities. A breakthrough was the development of stabilized finite element methods based on the subgrid scale approach (also referred to as the Variational Multiscale Method, VMM, or the Residual Based stabilization) (Hughes, 1995; Hughes *et al.*, 1998). The main idea of the method is to split the problem unknowns into a large scale that can be solved by the finite element mesh, and a fine scale (subgrid scale) whose effects onto the large scales can be somehow modeled. Applications of the method for the Navier-Stokes equations can be found e.g., in Codina (2002); Codina *et al.* (2007); Bazilevs *et al.* (2007), and respectively, for the wave equation and convected wave equation with applications to aeroacoustics in Codina (2008) and, Guasch and Codina (2007). Stabilization will be used throughout this work when necessary.

Concerning implementation aspects, all FEM formulations in this work have been programmed in an in-house finite element software based on Fortran language. The generation of the geometries with the corresponding finite element meshes and the postprocessing of the results coming from the finite element solver have been addressed using GID (`http://www.gidhome.com/`), 3ds Max and Matlab.

1.3 Aims of the thesis

The main purpose of this thesis is to address the numerical production of vowels and diphthongs within complex three-dimensional vocal tract geometries. To do so, a time-domain Finite Element strategy is selected to numerically solve the corresponding acoustic wave equations. Special attention is paid to the vocal tract acoustics modelling, which constitutes the main core of this thesis, as commented in the preceding sections.

The specific main goals of this work can be listed as follows:

1. Implement and validate a time-domain FEM formulation for vowel production.

2. Develop a numerical strategy for computing vocal tract impedances using the acoustic pressure as the only physical magnitude (i.e., without requiring the acoustic particle velocity).

3. Study the radiation effects that human head geometry simplifications produce on the generated vowel sounds.

4. Find a two-dimensional approach that provides a good balance in terms of computational time and voice quality.

5. Develop a time-domain FEM approach for diphthong production.

These objectives are organized as described in the following Section 1.4.

1.4 Thesis organization

This thesis is organized in two parts. The first one addresses the simulation of vowels and the second one deals with the generation of diphthongs. It is to be noted that the former has required most of the time of this thesis, since it has been found essential a good characterization of vowel simulation prior to attempting its dynamic counterpart, namely the diphthongs. Therefore, most chapters deal with vowel production.

In what concerns the production of vowels:

- In Chapter 2 the Finite Element formulation for vowel production is presented. Some numerical examples dealing with vocal tract acoustics analysis and vowel synthesis are reported.
- In Chapter 3 the FEM approach developed in Chapter 2 is validated against experiments performed with simplified mechanical replicas of the vocal tract.
- In Chapter 4 an approach to compute the radiation and input impedance of vocal tracts is derived. This is based on the adaptation of an experimental two-microphone transfer function method (TMTF) to the numerical framework.
- In Chapter 5 different strategies to account for acoustic radiation losses are reviewed and compared.
- In Chapter 6 geometry simplifications of a realistic human head are analyzed in terms of radiation effects by simulating vocal tract radiation impedances and transfer functions.
- In Chapter 7 a two-dimensional approach is presented as a good balance between voice quality and computational costs. This is based on a tuning process of 2D vocal tracts that allows mimicking to a large extent the acoustics of 3D vocal tracts.

With regard to the generation of diphthongs:

- In Chapter 8 a stabilized Finite Element formulation based on subgrid scales is derived to account for dynamic vocal tracts. Numerical examples include diphthong sounds generated with three-dimensional and two-dimensional vocal tracts.

Conclusions and future research lines are finally exposed in Chapter 9. Additionally, two appendices have been included for completeness. Appendix A contains analytical expressions for radiation load impedances, whereas Appendix B describes the waveform glottal models used in this work.

Part II

FEM generation of vowels

Chapter 2

Finite Element formulation for vowel production

In this chapter a Finite Element approach for vowel production will be presented. First, the problem statement for vocal tract acoustics will be derived, where the time domain wave equation for the acoustic pressure will be supplemented with appropriate boundary and initial conditions for the generation of vowels. Second, the resulting problem will be expressed in its variational form and then discretized in space using the Finite Element Method (FEM) and in time using finite differences, resulting in an explicit numerical scheme. Next, the methodology followed to perform a numerical simulation for vowel production will be explained. Finally, two applications of the proposed FEM approach will be presented. In the first one, a vocal tract acoustic analysis will be performed of the geometry simplifications of an MRI-based vocal tract for vowel /a/. In the second one, the same vowel will be synthesized so as to listen to the generated sound.

This chapter is partially based on the following work

- Marc Arnela and Oriol Guasch (2013), "Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method," Journal of the Acoustical Society of America, 133(6), pp. 4197–4209.

2.1 Introduction

As commented in the introductory Chapter 1, the Finite Element Method seems to be the most appropriate numerical method to solve the underlaying "equations of voice", since it allows us to properly consider intricate vocal tract geometries. On the other hand, working in the time domain rather than in the frequency one, seems to be a good option because most of the events produced during voice production are time-dependent (e.g., vocal folds oscillation, vocal tract movement, etc.). In the case of vowels, the linearized acoustic wave equation may suffice. Yet, two options for it can be found in the literature. The first one consists in solving the standard irreducible wave equation for the acoustic pressure (see e.g., Vampola *et al.*, 2008a, 2011; Arnela and Guasch, 2013; Arnela *et al.*, 2013; Arnela and Guasch, 2014b). The second one solves the mixed wave equation for the acoustic pressure and the acoustic particle velocity (see e.g, Takemoto *et al.*, 2010, 2013). Since the acoustic pressure is the magnitude of interest for voice production, it seems more appropriate to directly solve the irreducible wave equation. Moreover, this equation has the advantage that the standard Finite Element Galerkin discretization does not require stabilization, in contrast to the mixed wave equation (see e.g., Codina, 2008), numerical efforts usually being placed to the time discretization scheme.

In this chapter the time domain wave equation for the acoustic pressure will be solved using FEM. First, in Section 2.2 this equation will be supplemented with appropriate initial and boundary conditions for vowels. These include an inflow generated by the vocal folds, wall losses and free-field radiation conditions. The Perfectly Matcher Layer formulation described in Grote and Sim (2010) and Kaltenbacher *et al.* (2013) will be adopted for the latter. Then, in Section 2.3 the resulting problem will be numerically solved using FEM. As usual, the problem will be first expressed in its variational or weak form (Section 2.3.1). Then, it will be discretized following the standard Galerkin approach (Section 2.3.2). Finally, an explicit numerical scheme will be obtained by also discretizing in time using finite differences (Section 2.3.3). In the following Section 2.4, some details will be given concerning the general procedure followed to perform a FEM simulation for vowels. Finally, in Section 2.5 two classical applications in the voice production field will be addressed using the proposed numerical approach. In the first one (Section 2.5.1), vocal tract acoustic analysis will be performed. The most typical magnitude that describes the acoustic behavior of a vocal tract, the vocal tract transfer function, will be introduced. As examples, the geometry simplifications of an MRI-based vocal tract for vowel /a/ will be analyzed. In the second application (Section 2.5.2), FEM simulations using simplified vocal tracts will be performed to synthesize two examples corresponding to different geometries for vowel /a/. As a result, not only the acoustic pressure will be visualized, but also an audio file will be generated to listen to the produced vowel sound.

2.2 Problem statement

2.2.1 The acoustic wave equation

The mathematical problem to be solved by the finite element method will be next exposed. Let us denote by Ω the computational domain where acoustic waves propagate. Ω comprises the vocal tract and the outer region surrounding the head, where waves emanating from the mouth propagate in free field conditions. The boundary of Ω, $\partial\Omega$, can be split into four non-intersecting regions (see Fig. 2.1): Γ_{G} stands for the vocal tract cross-section at the vocal folds position, Γ_{W} corresponds to the vocal tract walls, Γ_{H} to the head contour and Γ_∞ is a fictitious non-reflecting boundary. The acoustic pressure which accounts for vowel generation can be obtained from the solution of the wave equation

$$\left(\partial_{tt}^2 - c_0^2\nabla^2\right) p = 0 \qquad \text{in } \Omega,\ t > 0, \tag{2.1a}$$

with boundary and initial conditions

$$\nabla p \cdot \boldsymbol{n} = g \qquad \text{on } \Gamma_{\mathrm{G}},\ t > 0, \tag{2.1b}$$

$$\nabla p \cdot \boldsymbol{n} = -\mu/c_0 \partial_t p \qquad \text{on } \Gamma_{\mathrm{W}},\ t > 0, \tag{2.1c}$$

$$\nabla p \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\mathrm{H}},\ t > 0, \tag{2.1d}$$

$$\nabla p \cdot \boldsymbol{n} = 1/c_0 \partial_t p \qquad \text{on } \Gamma_\infty,\ t > 0, \tag{2.1e}$$

$$p = 0,\ \partial_t p = 0 \qquad \text{in } \Omega,\ t = 0. \tag{2.1f}$$

In Eq. (2.1) c_0 stands for the speed of sound, $p(\boldsymbol{x}, t)$ is the acoustic pressure, $g(t)$ is a function related to the airflow generated by the vocal folds (glottal pulses), $\mu(\boldsymbol{x})$ is a friction coefficient for the losses at the vocal tract walls, $\boldsymbol{n}$ is the normal vector pointing outwards $\partial\Omega$ and $\partial_t \equiv \partial/\partial t$ denotes the partial time derivative.

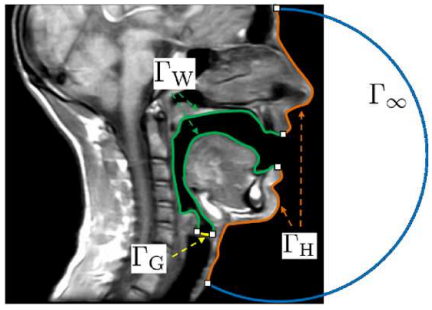


Fig. 2.1: A sketch of the computational domain Ω of Eq. (2.1) in text. Γ_G represents the glottal cross-sectional area, Γ_W the vocal tract walls, Γ_H the human head and Γ_∞ a non-reflecting boundary.

2.2.2 Glottal flow

The boundary condition in Eq. (2.1b) accounts for the volume velocity

$$Q_g(t) = u_g(t)A_g \tag{2.2}$$

generated by the vocal folds, where u_g stands for the inflow velocity and A_g for the area of the boundary Γ_{G}. The time derivative of the volume airflow $\partial_t Q_g$ is provided by most parametric glottal models such as the Rosenberg (Rosenberg, 1971) or LF (Fant *et al.*, 1985) ones (see also Appendix B). Then, it follows from the momentum equation that

$$\nabla p \cdot \boldsymbol{n} = -\rho_0 \partial_t u_g = -\rho_0 \frac{\partial_t Q_g}{A_g} \equiv g, \tag{2.3}$$

where ρ_0 designates the air density and function $g(t)$ in (2.1b) is defined in the last equality.

2.2.3 Wall losses

The subsequent boundary condition in Eq. (2.1c) considers losses at the vocal tract walls (locally reacting soft walls). It arises from the following reasoning. The impedance of the wall $Z(\boldsymbol{x}, \omega)$ relates the acoustic velocity at the wall $U(\boldsymbol{x}, \omega)$ and the acoustic pressure $P(\boldsymbol{x}, \omega)$,

$$U(\boldsymbol{x}, \omega) = \frac{P(\boldsymbol{x}, \omega)}{Z(\boldsymbol{x}, \omega)}. \tag{2.4}$$

The momentum equation in the frequency domain can then be written as ($\partial t \leftrightarrow j\omega$, with $j = \sqrt{-1}$)

$$\nabla P(\boldsymbol{x}, \omega) \cdot \boldsymbol{n} = -j\omega\rho_0 \frac{P(\boldsymbol{x}, \omega)}{Z(\boldsymbol{x}, \omega)}. \tag{2.5}$$

Assuming a frequency independent impedance, introducing the boundary admittance coefficient

$$\mu = \frac{\rho_0 c_0}{Z}, \tag{2.6}$$

replacing it in (2.5) and transforming the resulting expression back to the time domain, results in Eq. (2.1c). In what concerns the homogeneous Neumann boundary condition in Eq. (2.1d) it simply states that the surface of the head is modeled as a hard wall (i.e., $Z(\boldsymbol{x}, \omega) = \infty$). On the other hand, the boundary admittance coefficient can also be related with the common sound absorption coefficient α_n as (see e.g., Takemoto *et al.*, 2010)

$$\mu = \frac{1 - \sqrt{1 - \alpha_n}}{1 + \sqrt{1 - \alpha_n}}. \tag{2.7}$$

2.2.4 Free-field radiation: the Perfectly Matched Layer

Finally, Eq. (2.1e) is the well-known Sommerfeld radiation condition, which guarantees that the emitted waves propagate outwards to infinity. However, this condition is only optimal for waves impinging orthogonal onto Γ_∞. To avoid this problem and to perform simulations in a computational domain of reasonable size as well, Eq. (2.1e) is replaced with a Perfectly Matched Layer. PMLs were originally introduced by Berenger (1994) and are regions designed to absorb waves incident from any direction without producing reflection at its interface (see Fig. 2.2a).

The PML formulation of Grote and Sim (2010) has been adapted for our custom code, which solves Eq. (2.1) with the PML using FEM. Replacing the Sommerfeld radiation condition with the PML, the original acoustic wave equation (2.1) becomes modified to

$$\partial_{tt}^2 p - c_0^2 \nabla^2 p = \nabla \cdot \boldsymbol{\phi} - \alpha \partial_t p - \beta p - \gamma \psi \qquad \text{in } \Omega,\ t > 0, \tag{2.8a}$$

$$\partial_t \phi_i = -\xi_i \phi_i + c_0^2 a_i \partial_i p + c_0^2 b_i \partial_i \psi,\ \forall i = 1,2,3 \qquad \text{in } \Omega,\ t > 0, \tag{2.8b}$$

$$\partial_t \psi = p \qquad \text{in } \Omega,\ t > 0, \tag{2.8c}$$

with boundary and initial conditions

$$\nabla p \cdot \boldsymbol{n} = -\rho_0 / A_g \partial_t Q_g \equiv g \qquad \text{on } \Gamma_{\mathrm{G}},\ t > 0, \tag{2.8d}$$

$$\nabla p \cdot \boldsymbol{n} = -\mu / c_0 \partial_t p \qquad \text{on } \Gamma_{\mathrm{W}},\ t > 0, \tag{2.8e}$$

$$\nabla p \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\mathrm{H}},\ t > 0, \tag{2.8f}$$

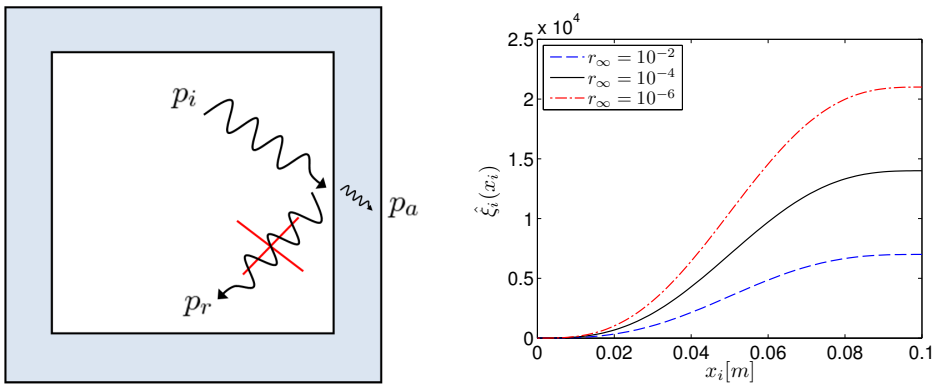
$$\nabla p \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_\infty,\ t > 0, \tag{2.8g}$$

$$p = 0,\ \partial_t p = 0 \qquad \text{in } \Omega, t = 0, \tag{2.8h}$$

$$\phi_i = 0,\ \partial_t \phi_i = 0,\ \forall i = 1,2,3 \qquad \text{in } \Omega, t = 0, \tag{2.8i}$$

$$\psi = 0,\ \partial_t \psi = 0 \qquad \text{in } \Omega, t = 0. \tag{2.8j}$$

Note that (2.8a) modifies the right hand side (r.h.s.) of (2.1a) with the inclusion of some extra terms involving the auxiliary functions ψ and $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)$. Moreover, four additional scalar equations (2.8b)-(2.8c) are needed for these additional functions. Although it may be thought that the above formulation could result in a high computational cost, it should be noticed that the additional variables are only non-null in the PML region, and that the computational domain can be much smaller than the one needed if the PML was absent. On the other hand, it has recently been observed that enhanced stability properties can be obtained when setting $b_i = 0$ in Eq. (2.8b) (Kaltenbacher *et al.*, 2013), allowing one to attain long time simulations. However, in this work all terms will be kept to have a greater accuracy, since the performed FEM simulations mainly involve short time events ($\sim$25 ms). In what concerns the coefficients



(a) (b)

Fig. 2.2: (a) Sketch of a computational domain using a Perfectly Matched Layer (shaded area), where an incident wave p_i reaches the PML interface and becomes absorbed inside it (p_a), without giving to any reflection (p_r). (b) Damping profile $\hat{\xi}_i$ for different reflection coefficients r_∞ at the PML boundary considering $L_i = 0.1$ m and $c = 350$ m/s.

α, β, γ, a_i, b_i, they depend on the damping profiles ξ_i and are given by

$$\alpha = \xi_1 + \xi_2 + \xi_3, \qquad a_1 = \xi_2 + \xi_3 - \xi_1, \qquad b_1 = \xi_2\xi_3, \tag{2.9a}$$

$$\beta = \xi_1\xi_2 + \xi_2\xi_3 + \xi_3\xi_1, \qquad a_2 = \xi_3 + \xi_1 - \xi_2, \qquad b_2 = \xi_3\xi_1, \tag{2.9b}$$

$$\gamma = \xi_1\xi_2\xi_3, \qquad a_3 = \xi_1 + \xi_2 - \xi_3, \qquad b_3 = \xi_1\xi_2. \tag{2.9c}$$

The damping profiles ξ_i are used to control the amount of absorption in the PML and many options exist for them. Following Grote and Sim (2010), use has been made of

$$\xi_i(x_i) = \hat{\xi}_i \left[\frac{|x_i - l_i|}{L_i} - \frac{\sin\left(\frac{2\pi|x_i - l_i|}{L_i}\right)}{2\pi} \right] \tag{2.10}$$

for $l_i \leq |x_i| \leq l_i + L_i$. $\hat{\xi}_i$ is a constant accounting for the damping effect in the i-th direction, l_i is the i-th coordinate of the PML layer and L_i the thickness of the PML region in the i-th direction (see Fig. 2.2b). The constant $\hat{\xi}_i$ depends on the discretization and thickness of the layer and can be computed as

$$\hat{\xi}_i = \frac{c_0}{L_i} \log\left(\frac{1}{r_\infty}\right), \tag{2.11}$$

with r_∞ standing for the relative reflection at the boundary of the PML. The PML boundary can be truncated using either a Dirichlet or a Neumann homogeneous condition (the latter has been our choice, see Eq. (2.8g)). On the other hand, notice that for $l_i \leq |x_i|$, $\xi_i(x_i) = 0$, i.e., outside the PML, the damping profiles ξ_i vanish. In this case, the modified PML wave equation (2.8a) reduces to the standard wave equation in (2.1a).

2.3 Finite element approach

2.3.1 Variational formulation

The set of partial differential equations (2.8a)-(2.8c) supplemented with boundary and initial conditions (2.8d)-(2.8j) constitutes the problem to be solved. If a FEM approach is to be used to find a numerical solution for (2.8), first the problem has to be set in its weak or variational form. Let us introduce some additional notation to do so, without getting into much deep mathematical details. The space of square integrable functions in Ω will be denoted as usual by $L^2(\Omega)$, whereas the space of functions in $L^2(\Omega)$ whose spatial derivatives are also square integrable will be denoted by $H^1(\Omega)$. V_p will denote the space of functions for the unknowns that belong to $H^1(\Omega)$ for every time instant t and satisfy the problem essential boundary conditions, and V_q will stand for those functions in H^1 whose values vanish on the boundaries. To shorten the notation, the integral of the product of any two functions f, g in Ω will be written as

$$(f, g) := \int_\Omega f g d\Omega, \tag{2.12}$$

whilst integrals over boundaries will be explicitly indicated e.g., $(f, g)_{\Gamma_G}$.

Once established the above space functional framework, we can proceed to find the variational formulation for problem (2.8). As usual, this is done by first multiplying (2.8) by test functions q, v_i, w (q for the pressure, v_i for the first auxiliary functions ϕ_i, and w for the second auxiliary function ψ) and integrating over the computational domain Ω. Applying the divergence theorem and making use of boundary conditions, the weak form of the problem is obtained, which consists in finding p, ϕ_i and $\psi \in V_p$ such that

$$\begin{aligned}\left(q, \partial_{tt}^2 p\right) + c_0 \left(q, \mu \partial_t p\right)_{\Gamma_{\mathrm{W}}} + c_0^2 \left(\nabla q, \nabla p\right) &= c_0^2 \left(q, g\right)_{\Gamma_{\mathrm{G}}} \\ &+ \sum_{i=1}^{3} \left(q, \partial_i \phi_i\right) - \left(q, \alpha \partial_t p\right) - \left(q, \beta p\right) - \left(q, \gamma \psi\right), \end{aligned} \tag{2.13a}$$

$$\left(v_i, \partial_t \phi_i\right) = -\left(v_i, \xi_i \phi_i\right) + c_0^2 \left(v_i, a_i \partial_i p\right) + c_0^2 (v_i, b_i \partial_i \psi), \ \forall i = 1, 2, 3, \tag{2.13b}$$

$$\left(w, \partial_t \psi\right) = \left(w, p\right), \tag{2.13c}$$

in Ω, $t > 0$, with initial conditions

$$\left(q, p\right) = 0, \ \ \left(q, \partial_t p\right) = 0, \tag{2.13d}$$

$$\left(v_i, \phi_i\right) = 0, \ \ \left(v_i, \partial_t \phi_i\right) = 0, \ \forall i = 1, 2, 3, \tag{2.13e}$$

$$\left(w, \psi\right) = 0, \ \ \left(w, \partial_t \psi\right) = 0, \tag{2.13f}$$

in $\Omega, t = 0$, for all $q, v_i, w \in V_q$.

2.3.2 Spatial discretization: semi-discrete formulation

To find a numerical solution to Eq. (2.13), this has to be discretized in both, space and time. Let us first discretize it in space using a FEM approach. Given a finite element partition of Ω with n_{el} elements and n_p nodes and the finite dimensional subspaces $V_{p,h} \subset V_p$ and $V_{q,h} \subset V_q$, the Galerkin finite element approach to problem (2.13) aims at finding p_h, ϕ_{ih} and $\psi_h \in V_{p,h}$ such that

$$\begin{aligned}\left(q_h, \partial_{tt}^2 p_h\right) + c_0 \left(q_h, \mu \partial_t p_h\right)_{\Gamma_{\mathrm{W}}} + c_0^2 \left(\nabla q_h, \nabla p_h\right) &= c_0^2 \left(q_h, g\right)_{\Gamma_{\mathrm{G}}} \\ + \sum_{i=1}^{3} \left(q_h, \partial_i \phi_{ih}\right) - \left(q_h, \alpha_h \partial_t p_h\right) - \left(q_h, \beta_h p_h\right) - \left(q_h, \gamma_h \psi_h\right), & \quad (2.14a)\end{aligned}$$

$$\left(v_{ih}, \partial_t \phi_{ih}\right) = -\left(v_{ih}, \xi_{ih} \phi_{ih}\right) + c_0^2 \left(v_{ih}, a_{ih} \partial_i p_h\right) + c_0^2 (v_{ih}, b_{ih} \partial_i \psi_h), \ \forall i = 1, 2, 3, \quad (2.14b)$$

$$\left(w_h, \partial_t \psi_h\right) = \left(w_h, p_h\right), \quad (2.14c)$$

in $\Omega,\ t > 0$, with initial conditions

$$\left(q_h, p_h\right) = 0, \ \ \left(q_h, \partial_t p_h\right) = 0, \quad (2.14d)$$

$$\left(v_{ih}, \phi_{ih}\right) = 0, \ \ \left(v_{ih}, \partial_t \phi_{ih}\right) = 0, \ \forall i = 1, 2, 3, \quad (2.14e)$$

$$\left(w_h, \psi_h\right) = 0, \ \ \left(w_h, \partial_t \psi_h\right) = 0, \quad (2.14f)$$

in $\Omega, t = 0$, for all $q_h, v_{ih}, w_h \in V_{q,h}$. The subscripts h in (2.14) denote the discrete versions of the corresponding continuous variables. Expanding the unknowns p_h, ϕ_{ih}, ψ_h and the test functions q_h, v_{ih} and w_h in terms of shape functions $N(\boldsymbol{x})$ (piecewise linear shape functions will be used in this work) and nodal values $P^b = p_h(\boldsymbol{x}^b, t)$, $\Phi_i^b = \phi_{ih}(\boldsymbol{x}^b, t)$, $\Psi^b = \psi_h(\boldsymbol{x}^b, t)$, $Q^a = q_h(\boldsymbol{x}^a)$, $V_i^a = v_{ih}(\boldsymbol{x}^a)$ and $W^a = w_h(\boldsymbol{x}^a)$ we get

$$p_h = \sum_{b=1}^{n_p} N^b P^b, \qquad \phi_{ih} = \sum_{b=1}^{n_p} N^b \Phi_i^b, \qquad \psi_h = \sum_{b=1}^{n_p} N^b \Psi^b, \quad (2.15a)$$

$$q_h = \sum_{a=1}^{n_p} N^a Q^a, \qquad v_{ih} = \sum_{a=1}^{n_p} N^a V_i^a, \qquad w_h = \sum_{a=1}^{n_p} N^a W^a. \quad (2.15b)$$

Substituting (2.15) in the spatial discretized scheme (2.14) results in the following time evolving algebraic matrix system

$$\begin{aligned}\boldsymbol{M}\ddot{\boldsymbol{P}} + c_0 \boldsymbol{B}\dot{\boldsymbol{P}} + c_0^2 \boldsymbol{K}\boldsymbol{P} &= c_0^2 \boldsymbol{L} \\ + \sum_{i=1}^{3} \boldsymbol{B}_i \boldsymbol{\Phi}_i - \boldsymbol{M}_\alpha \dot{\boldsymbol{P}} - \boldsymbol{M}_\beta \boldsymbol{P} - \boldsymbol{M}_\gamma \boldsymbol{\Psi}, & \quad (2.16a)\end{aligned}$$

$$\boldsymbol{M}\dot{\boldsymbol{\Phi}}_i = -\boldsymbol{M}_{\xi_i} \boldsymbol{\Phi}_i + c_0^2 \boldsymbol{B}_{i,a_i} \boldsymbol{P} + c_0^2 \boldsymbol{B}_{i,b_i} \boldsymbol{\Psi}, \ \forall i = 1, 2, 3, \quad (2.16b)$$

$$\dot{\boldsymbol{\Psi}} = \boldsymbol{P}, \quad (2.16c)$$

in $\Omega,\ t > 0$, with initial conditions

$$\boldsymbol{P} = 0,\ \dot{\boldsymbol{P}} = 0, \tag{2.16d}$$

$$\boldsymbol{\Phi}_i = 0,\ \dot{\boldsymbol{\Phi}}_i = 0,\ \forall i = 1, 2, 3, \tag{2.16e}$$

$$\boldsymbol{\Psi} = 0,\ \dot{\boldsymbol{\Psi}} = 0, \tag{2.16f}$$

in $\Omega, t = 0$. In Eq. (2.16), $\boldsymbol{P}$, $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Psi}$ stand for the vectors of nodal values that respectively correspond to the pressure and auxiliary functions (e.g., $\boldsymbol{P} = (P^1 \cdots P^{n_p})^\top$), whereas the remaining matrix and vector entries are given by

$$M^{ab} = \left(N^a, N^b\right), \qquad M^{ab}_\alpha = \left(N^a, \alpha_h N^b\right), \tag{2.17a}$$

$$M^{ab}_\beta = \left(N^a, \beta_h N^b\right), \qquad M^{ab}_\gamma = \left(N^a, \gamma_h N^b\right), \tag{2.17b}$$

$$M^{ab}_{\xi_i} = \left(N^a, \xi_{ih} N^b\right), \qquad B^{ab} = \left(N^a, \mu N^b\right)_{\Gamma_\mathrm{W}}, \tag{2.17c}$$

$$B^{ab}_{i,a_i} = \left(N^a, a_{ih}\partial_i N^b\right), \qquad B^{ab}_i = \left(N^a, \partial_i N^b\right), \tag{2.17d}$$

$$B^{ab}_{i,b_i} = \left(N^a, b_{ih}\partial_i N^b\right), \qquad K^{ab} = \left(\nabla N^a, \nabla N^b\right), \tag{2.17e}$$

$$L^a = \left(N^a, g\right)_{\Gamma_\mathrm{G}}. \tag{2.17f}$$

As usual, the domain integrals in (2.17) are to be understood as the summation of integrals over elements Ω_e, i.e., $(\cdot,\cdot)_\Omega = \sum_{e=1}^{n_{el}} (\cdot,\cdot)_{\Omega_e}$.

2.3.3 Time discretization: final numerical scheme

Let us next proceed to the time discretization of (2.16). A finite difference approach has been used to do so. Considering a constant time step ($\Delta t = t^{n+1} - t^n$) discretization of the time interval $(0, T)$ into $0 < t^1 < \cdots < t^{n-1} < t^n < t^{n+1} < \cdots < t^N \equiv T$. A second order finite difference central scheme has then been implemented for the pressure time derivatives, whilst a first order central scheme has been used for the time derivatives of the auxiliary variables in the PML region. This is so because first order schemes are known to introduce stronger numerical dissipation, which in this case is advantageous to help absorbing the waves crossing the PML. Inserting these time derivative approximations in (2.16) yields

$$\begin{aligned}\boldsymbol{M}\frac{\boldsymbol{P}^{n+1} - 2\boldsymbol{P}^n + \boldsymbol{P}^{n-1}}{\Delta t^2} + c_0\boldsymbol{B}\frac{\boldsymbol{P}^{n+1} - \boldsymbol{P}^{n-1}}{2\Delta t} + c_0^2\boldsymbol{K}\boldsymbol{P}^n = c_0^2\boldsymbol{L}^n \\ + \sum_{i=1}^{3}\boldsymbol{B}_i\boldsymbol{\Phi}_i^n - \boldsymbol{M}_\alpha\frac{\boldsymbol{P}^{n+1} - \boldsymbol{P}^{n-1}}{2\Delta t} - \boldsymbol{M}_\beta\boldsymbol{P}^n - \boldsymbol{M}_\gamma\frac{\boldsymbol{\Psi}^{n+1/2} + \boldsymbol{\Psi}^{n-1/2}}{2},\end{aligned} \tag{2.18a}$$

$$\boldsymbol{M}\frac{\boldsymbol{\Phi}_i^{n+1} - \boldsymbol{\Phi}_i^n}{\Delta t} = -\boldsymbol{M}_{\xi_i}\frac{\boldsymbol{\Phi}_i^{n+1} + \boldsymbol{\Phi}_i^n}{2} + c_0^2\boldsymbol{B}_{i,a_i}\frac{\boldsymbol{P}^{n+1} + \boldsymbol{P}^n}{2} + c_0^2\boldsymbol{B}_{i,b_i}\boldsymbol{\Psi}^{n+1/2}, \tag{2.18b}$$

$$\frac{\boldsymbol{\Psi}^{n+1/2} - \boldsymbol{\Psi}^{n-1/2}}{\Delta t} = \boldsymbol{P}^n, \tag{2.18c}$$

with initial conditions

$$\boldsymbol{P}^0 = 0, \ \boldsymbol{P}^1 = 0, \tag{2.18d}$$

$$\boldsymbol{\Phi}_i^0 = 0, \ \boldsymbol{\Phi}_i^1 = 0, \ \forall i = 1, 2, 3, \tag{2.18e}$$

$$\boldsymbol{\Psi}^{-1/2} = 0, \ \boldsymbol{\Psi}^{1/2} = 0. \tag{2.18f}$$

Note that (2.18) corresponds to a purely explicit scheme, where all the unknowns can be calculated at time step $n+1$ from values already known from previous steps. In contrast to the implicit ones, explicit schemes avoid matrix inversion, although at the price of requiring very small time steps to fulfill a restrictive Courant-Friedrich-Levy condition (the smaller the element size, the smaller the time step $\Delta t = 1/f_s$ has to be).

Solving (2.18) at time step $t = n + 1$ involves

1. Use (2.18c) to compute $\boldsymbol{\Psi}^{n+1/2}$
2. Insert $\boldsymbol{\Psi}^{n+1/2}$ into (2.18a) and compute $\boldsymbol{P}^{n+1}$
3. Update $\boldsymbol{\Phi}_i^{n+1}$ using (2.18b), $\boldsymbol{\Psi}^{n+1/2}$ and $\boldsymbol{P}^{n+1}$

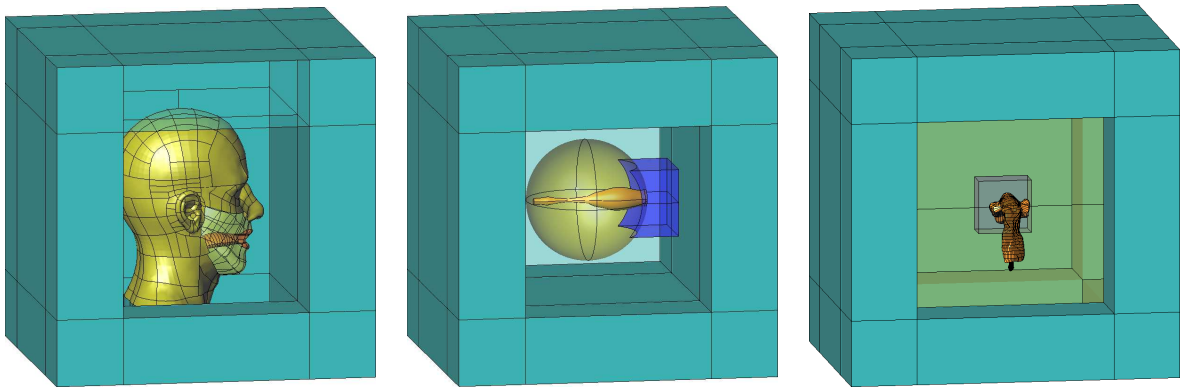
To fasten all computations, matrix inversion is avoided as usual by means of a lumped approximation for all mass matrices (2.17a)-(2.17c) (Hughes, 2000). The system of equations (2.18) constitutes the final numerical scheme that has been used for all the computational simulations with regard to vowel generation.

2.4 General description of the FE simulations

The following computational features have been used to compute the acoustic pressure for the numerical examples. In case of any modification of the below parameters for a specific simulation, this will be explicitly indicated in the corresponding section.

The values $c_0 = 350$ m/s and $\rho_0 = 1.14$ kg/m^3 have respectively been chosen for the speed of sound and air density in Eq. (2.1). Concerning boundary conditions, a wideband impulse has been used for the volume velocity $Q_g(t)$ in Eq. (2.3), consisting of a Gaussian pulse of the type (Takemoto *et al.*, 2010)

$$gp(n) = e^{[(\Delta t \ n - T_{gp})0.29T_{gp}]^2} [\mathrm{m}^3/\mathrm{s}], \tag{2.19}$$



(a) Realistic head (b) Spherical baffle (c) Flat baffle

Fig. 2.3: Computational domain for vowel production simulation where a vocal tract is set on different radiation elements. The most outer volumes correspond to the PML, which are surrounding the free-field radiation space.

with $T_{gp} = 0.646/f_0$ and $f_0 = 10$ kHz. To avoid numerical errors beyond the maximum frequency of interest ($f_{max} = 10$ kHz), this pulse has been filtered using a low-pass filter with cutoff frequency 10 kHz. For the boundary admittance coefficient at the vocal tract walls the value $\mu_w = 0.005$ has been used, which corresponds to the wall impedance of the vocal tract tissue $Z_w = 83666$ kg/m^{2}s (see e.g., Švancara and Horáček, 2006).

With regard to the computational domain (see Fig. 2.3), the vocal tracts or duct systems are set either on flat baffles, spherical baffles or realistic human heads, depending on the required degree of accuracy of the radiation model (this will be analyzed in more detail in Chapter 5 and Chapter 6). These get surrounded with corresponding volumes to allow sound waves emanate from the mouth exit and, in turn, these volumes are enclosed with a Perfectly Matched Layer, 0.1 m width, to absorb any incident wave. The PML has been configured to get a reflection coefficient of $r_\infty = 10^{-4}$. The resulting computational domains have been meshed using tetrahedral elements. The mesh size h ranges from $h = 0.001$ m within the vocal tract to $h = 0.002$ m in the mouth immediate outer region, $h = 0.005$ m in the most outer volume, and $h = 0.0075$ m in the PML.

Finally, Eq. (2.1) with a PML included has been solved using the explicit finite element numerical scheme described in Eq. (2.18). A sampling rate of $f_s = 1/\Delta t = 2000$ kHz is used to fulfil a stability condition of the Courant-Friedrich-Levy type. In some simulations this value has been increased to guarantee the stability condition. A time interval of total duration $T = 25$ ms has been simulated.

2.5 Numerical examples

2.5.1 Vocal tract acoustics analysis

One of the typical applications of Finite Element simulations for vowel production consists in analyzing the vocal tract acoustic behavior. The latter can be characterized by computing the so called vocal tract transfer functions $H(f)$, which can be obtained from the ratio

$$H(f) = \frac{P_o(f)}{Q_g(f)}, \tag{2.20}$$

where $P_o(f)$ and $Q_g(f)$ respectively stand for the Fourier Transform of the acoustic pressure $p_o(t)$ captured at the exit of the vocal tract and the volume velocity $Q_g(t)$ introduced at the glottal cross-section. In this case the latter does not correspond to the typical train of glottal pulses generated by a glottal source model but rather to a Gaussian pulse to excite all frequencies. In particular we use that described in Eq. (2.19).

To exemplify this type of analysis the following example has been chosen. In the classical approximation of 1D models the vocal tract geometry is represented by the so called area functions (see e.g., Story, 2008; Story *et al.*, 1996), which describe the area of the vocal tract cross-sections along its midline. The process to compute these area functions can be summarized as follows. A volumetric representation of the vocal tract is first generated from a set of magnetic resonances images (MRI) captured in different views, typically in the sagittal, coronal and/or transverse planes. This vocal tract is then discretized in a finite set of cross-sections perpendicular to the vocal tract midline, from which the area of each cross-section is extracted. These values constitute the area function for the analyzed sound. In Fig. 2.4 this process is illustrated for vowel /a/. Figure 2.4a represents the reconstructed MRI model (from Palo *et al.*, 2012, original geometry available online at `http://speech.math.aalto.fi/data.html`), which contains not only the main conduct of the vocal tract but also different side branches like the piriform fossae and valleculae. In Fig. 2.4b the side branches have been removed and the MRI model has been discretized with 40 cross-sections (see e.g., Story *et al.*, 1996, where use is made of $40 - 46$ cross-sections). The vocal tract shape is approximated as having circular cross-sections in Fig. 2.4c, and finally the vocal tract is next unbent in Fig. 2.4d. The latter corresponds to the typical three-dimensional vocal tract that can be generated from area functions.

In this example we have computed the vocal tract transfer functions of each one of the vocal tracts shown in Fig. 2.4 to analyze the effects of vocal tract simplifications. In order to perform a numerical simulation these vocal tracts have been set in rigid flat baffle of dimensions 0.3×0.3 m, to be interpreted as a rough approximation of the human head.

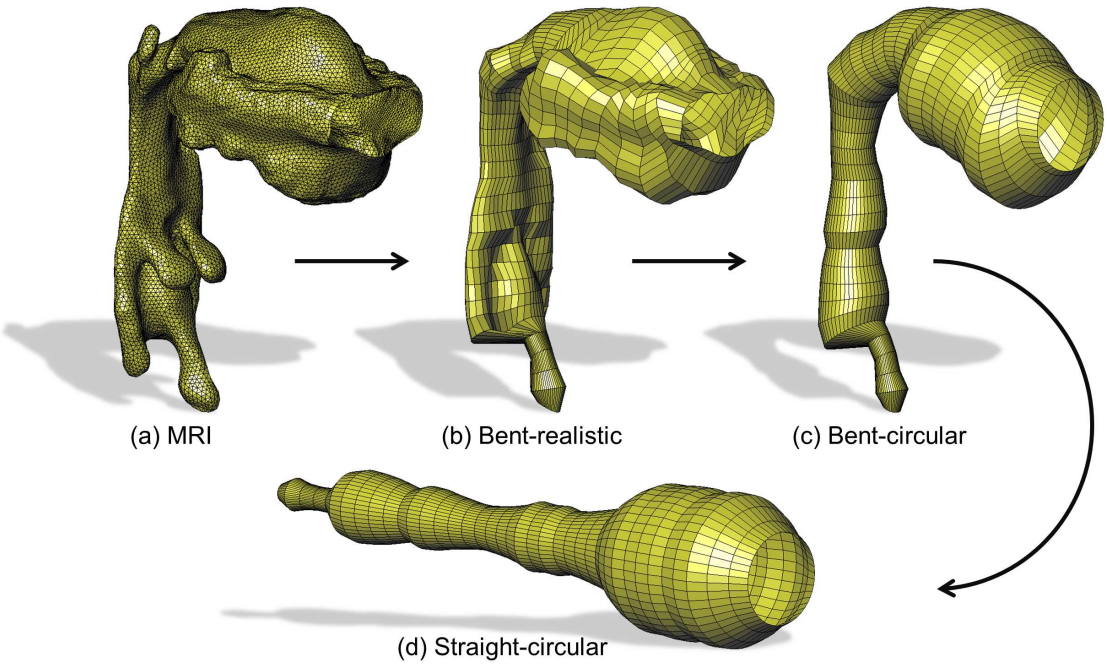


Fig. 2.4: Vocal tracts for vowel /a/ illustrating the possible steps followed to extract a one-dimensional area function. (a) Volume geometry reconstructed from MRI, (b) bent vocal tract with realistic cross-sections, (c) bent vocal tract with circular cross-sections and (d) straight vocal tract with circular cross-sections. The vocal tract simplifications have been constructed from a joint collaboration with Saeed Dabbaghchian and Olov Engwall.

This baffle constitutes one side of a rectangular volume that allows sound waves emanate from the vocal tract, which is then surrounded with a PML to absorb outgoing radiated waves. More numerical details can be found in Section 2.4.

The computed VTTFs are shown in Fig. 2.5. Resonances (known as formants in the voice and speech community) and antiresonances can be observed in the VTTFs. The location of the first two formants determines the produced vowel (vowel identification), while higher formants and antiresonances are responsible for other vowel features such as timber (vowel naturalness). If we first focus in the MRI model, we can observe that below $\sim$ 3 kHz only resonances appear. Beyond this value antiresonances are also generated, which can be attributed to the side branches and to high-order modes present within the main conduct of the vocal tract. Comparing the MRI case with any of the simplifications below $\sim$ 3 kHz, we can observe that the formant locations are quite similar between cases, since in this regime the plane wave assumption holds and therefore the shape of the vocal tract practically does not influence its acoustic behavior. Note that these include the first two formants and therefore one will identify the same

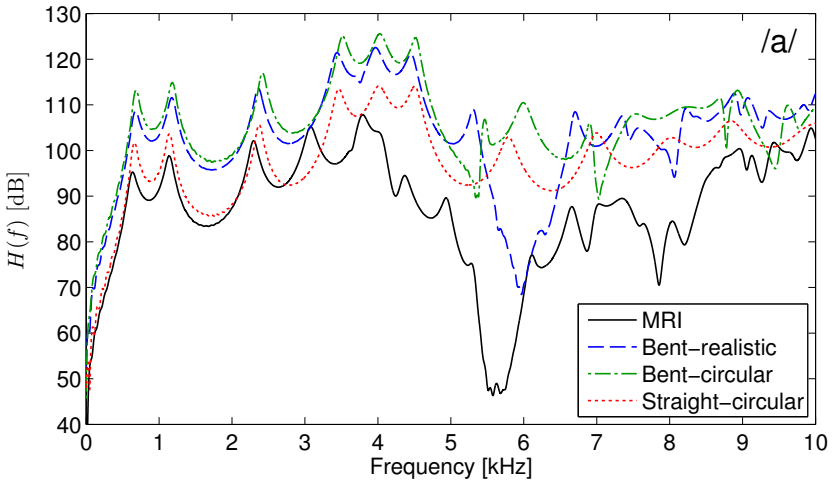


Fig. 2.5: Vocal tract transfer functions for vowel /a/ computed in the FEM numerical simulations using the vocal tract geometries in Fig. 2.4.

vowel sound for all cases, as one could expect. However, beyond this limit significant differences are found. Between $\sim$ 3 kHz and $\sim$ 5 kHz none of the simplifications match with the MRI case. The antiresonances of the MRI case that are not appearing in the other configurations may be produced by the side branches, which are not included in the simplifications (the effect of side branches such as the piriform fossae and valleculae is studied for instance in Takemoto *et al.*, 2010, 2013). Nonetheless, in this frequency range the formant frequencies should have a similar value independently of the case, since it seems that we are still in the plane wave propagation regime. Note for instance that the fourth formant does not significantly change between the three simplifications. However, some differences between the formants in the MRI case and those in the simplifications can still be appreciated. On the one hand, they can be attributed to the fact that the vocal tract has not probably been discretized with enough cross-sections, which produces smooth transitions of the vocal tract geometry instead of abrupt changes (e.g. the larynx opening into the pharynx, see Fig. 2.4). On the other hand, the computed midline from which the cross-sections are determined may not be accurate enough as several strategies can be found in literature (see for instance Story *et al.*, 1996; Takemoto *et al.*, 2006), none of them being determinant. Beyond 5 kHz, the vocal tract acoustic behavior becomes more intricate. Large differences can be observed between all cases. However, the case with realistic cross-sections still present some resonances and antiresonances that are also produced in the MRI geometry (see e.g. the deep antiresonance close to 6 kHz). Since this case does not contain side branches, these antiresonances could be attributed to transverse modes. Transforming the realistic shape of this case to a circular one produces a large difference in this high frequency range, although some small antiresonances can be

appreciated which can be attributed to the vocal tract bending. However, if we then make this vocal tract straight, although high-order modes can be excited their appearance in the vocal tract transfer functions is mitigated due to the radial symmetry of the geometry and therefore, only resonances are produced.

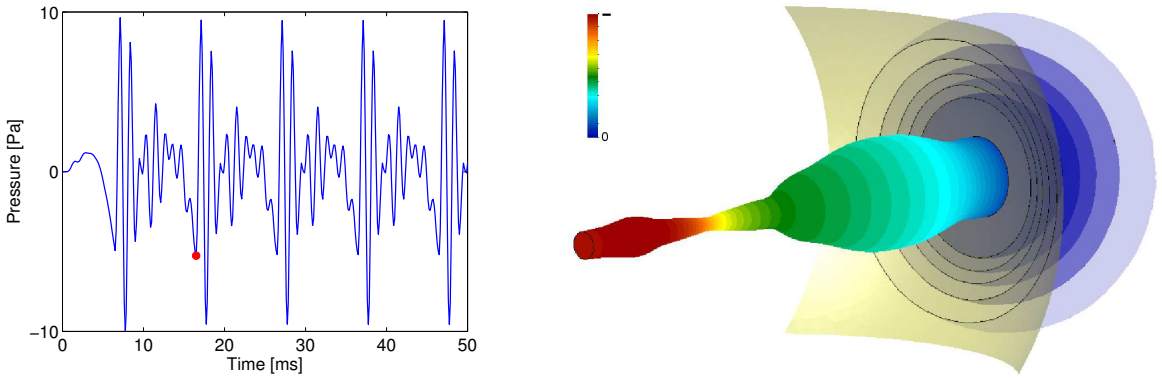
As observed, simplified vocal tract geometries do not accurately reproduce the resonances and antiresonances of the vocal tract, although in the low frequency region where plane wave propagation predominates they do it with a certain confidence. Hence the perceived vowel will be the same for a listener, although a poorer quality could be expected specially when a straight vocal tract with circular cross-sections is used. One should therefore consider, whenever is it possible, full detail vocal tracts to obtain a high voice quality. However, this is not always possible, since there are still very few vocal tract geometry databases available in the literature (e.g., Palo *et al.*, 2012; Aalto *et al.*, 2014) and moreover, they do not cover all sounds. In addition, the interpolation of static MRI geometries, or even the acquisition of dynamic MRI geometries, to generate for instance a diphthong sound is still a challenging problem. The alternative then consists in working with simplified vocal tract geometries, which can be generated using the large database of area functions available in the literature (e.g., Story *et al.*, 1996; Takemoto *et al.*, 2006; Story, 2008). Moreover, they provide a good simulation reproducibility given their generation simplicity and allow us to easily build dynamic geometries by interpolating for instance, the area functions. However, the vocal tract shape is not usually provided so one has to use for instance circular cross-sections. Through this work three-dimensional vocal tracts generated from area functions will be constructed.

2.5.2 Vowel synthesis

Some numerical examples dealing with vowel synthesis are presented in this section. In the first example the production of vowel /a/ is simulated. To do so, a simplified vocal tract for this vowel is built using the corresponding area function in Story (2008). These area functions are converted to circular cross-sections along the midline of the vocal tract which are then linearly interpolated to connect them. This vocal tract is set on a spherical baffle of radius 0.09 m, which can be interpreted as an approximation of the human head. The spherical surface is surrounded by a volume that allows emanating waves from the mouth to propagate towards infinity. A free-field radiation simulation is then ran using the computational features described in Section 2.4, but with the following modifications. To deal with a finite computational domain, a PML of width 0.2 m is used to absorb outward propagating waves. A train of glottal pulses with a fundamental frequency of $F_0 = 100$ Hz is introduced at the glottal area. These are generated using a LF glottal

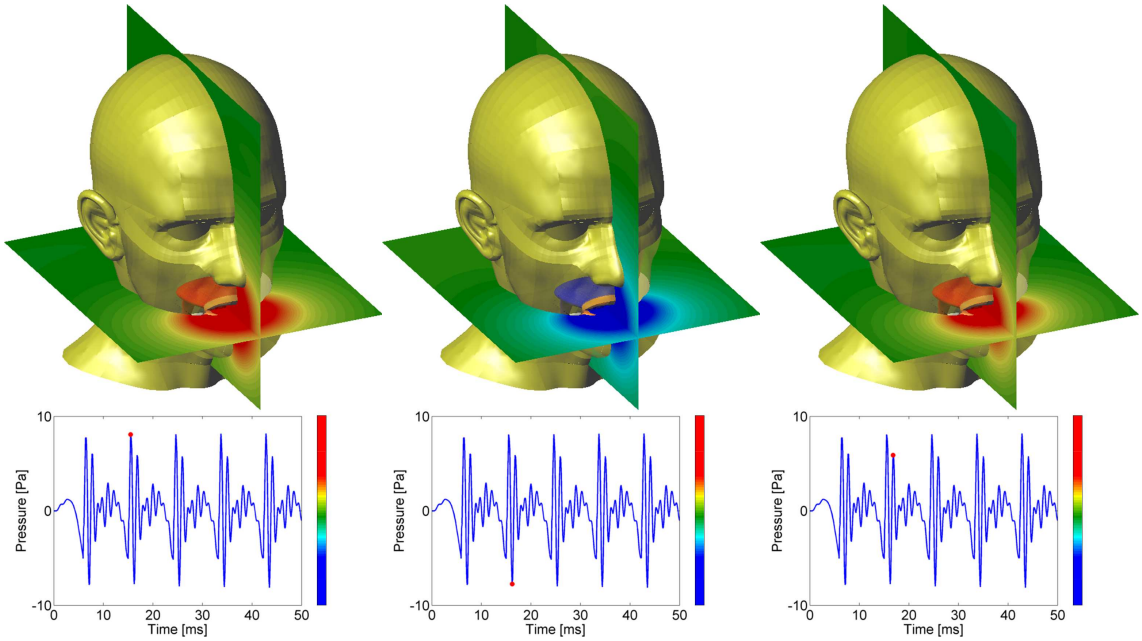
model (Fant *et al.*, 1985, see also Appendix B.2). A boundary admittance coefficient of $\mu = 0.005$ is set to account for losses at the walls. The computational domains are then meshed using tetrahedral elements. The mesh size h ranges approximately from $h \approx 0.001$ m inside the vocal tract to $h \approx 0.01$ m at the outer PML region. With regards to time evolution, the time discrete scheme (2.18) is used with a sampling rate of $f_s = 1/\Delta t = 1600$ kHz. The time evolution of the acoustic pressure is tracked at a node located 3 cm outside the mouth (see Fig. 2.6a). This sequence can be easily converted to an audio file to listen to the synthesized vowel. A snapshot of the numerical simulation at time instant $t \simeq 16$ ms (corresponding to the dot in the waveform of Fig. 2.6a) is presented in Fig. 2.6b, where it can be appreciated how plane front waves emanating from the vocal tract transform to spherical front waves, which propagate outwards.

A second example is also performed for the production of vowel /a/. However, in this case we consider a more complex geometry instead. The vocal tracts are generated using the same area functions from Story (2008), but they have been artificially bent and their cross-section shapes have been made elliptical. The resulting vocal tracts are set on a realistic human head. More details on the construction of this geometry are provided in Chapter 6. The features described in Section 2.4 are followed to perform a numerical simulation, but using a LF model to generate glottal pulses with a fundamental frequency of $F_0 = 110$ Hz. In Fig. 2.7 snapshots of the computed acoustic pressure are presented at different time instants. In each subfigure the vowel waveform can be appreciated with a dot denoting the time instant. This signal is tracked just in front of the mouth exit and converted to an audio file.

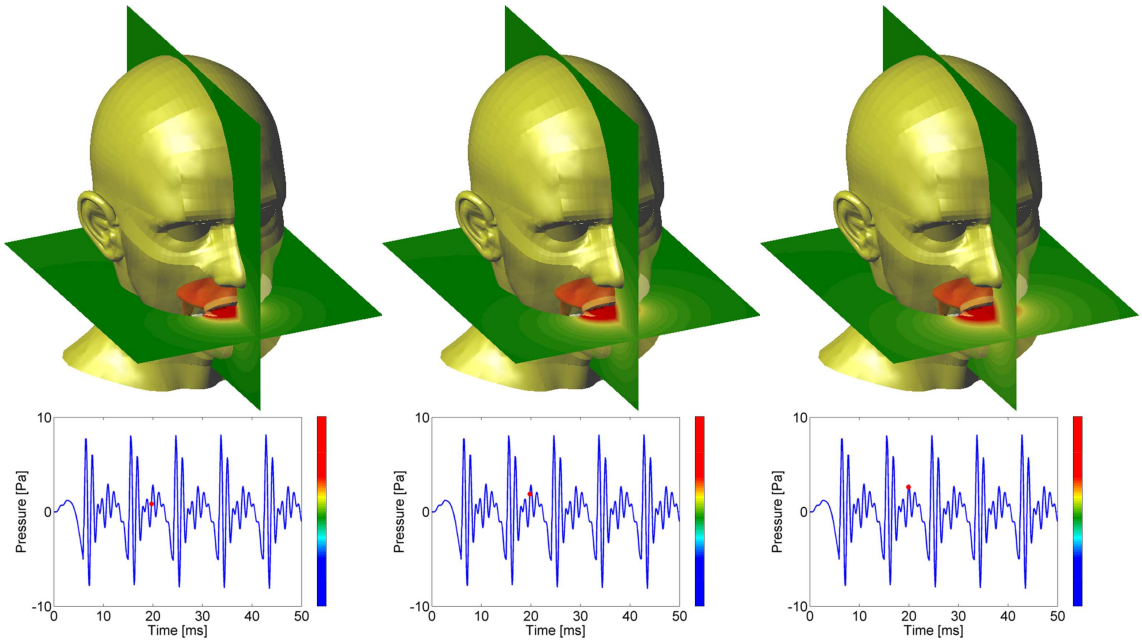


(a) (b)

Fig. 2.6: Example #1 of vowel synthesis for vowel /a/. (a) Waveform. (b) Wave propagation inside the vocal tract and emanating from mouth aperture at time instant $t = 13.3$ ms. Iso-pressure surfaces are used to represent the spherical waves emanating from the vocal tract. Note that the color scale is for negative values.



(a) $t = 15.6$ ms (b) $t = 16.3$ ms (c) $t = 16.9$ ms



(d) $t = 19.8$ ms (e) $t = 19.9$ ms (f) $t = 20.0$ ms

Fig. 2.7: Example #2 of vowel synthesis for vowel /a/. Snapshots at different time instants of the acoustic pressure evolution. Two cuts in the midsagittal and midtransverse planes are performed to better observe the acoustic waves outside the vocal tract. In each subfigure the resulting waveform tracked in front of the mouth aperture is also shown with a red dot denoting the corresponding time instant.

2.6 Conclusions

The wave equation for the acoustic pressure has been supplemented with appropriate initial and boundary conditions for vowel generation, which include the introduction of a volume velocity generated by the vocal folds at the glottal cross-section, a boundary admittance coefficient to account for wall losses and a PML to deal with free-field radiation. The resulting problem has then been numerically solved, using the FEM approach for the spatial discretization and finite differences for the discretization in time. This has yielded an explicit time scheme. Finally, two typical applications of numerical simulations of the vocal tract have been presented, showing a good performance of the presented formulation.

The first application has dealt with the analysis of vocal tract acoustics. This has been exemplified by studying the acoustic effects that geometry simplifications of an MRI-based vocal tract for vowel /a/ produce on the vocal tract transfer function. As one could expect, the spectra has not changed in the low frequency region (below 3 kHz), which guarantees that the same vowel sound is produced. However, between 3 kHz and 5 kHz some differences have been appreciated although the plane wave assumption still holds. These have been attributed to a poor spatial discretization of the MRI geometry in terms of cross-sections. In what concerns the high frequency region beyond 5 kHz, at first sight none of the simplifications has been able to correctly reproduce the behavior of the MRI geometry. A more detailed study is still necessary to verify some of the conclusions that have been reached at this preliminary stage. This will be performed in a near future. Despite of these clear limitations, simplifications of the vocal tract geometry are very useful for voice production simulation. For instance, they can be easily constructed from the large database of area functions available in the literature, and some processes such as generating dynamic vocal tracts become less intricate. Therefore, since the main goal of this work is to develop a framework to simulate the acoustics of the vocal tract, rather than to construct a complete voice simulator, in the following sections of this thesis use will be made of these simplified vocal tracts. It should be noted however that the presented approach is general enough so as to consider any more intricate geometry.

In what concerns the second application, two numerical examples have been shown dealing with the synthesis of vowel /a/. After introducing a train of pulses generated by a Rosenberg model or a LF model at the glottal cross-section, the acoustic pressure has been tracked at a node close to the mouth aperture and then transformed to an audio file so as to listen to the generated sound. Moreover, acoustic wave propagation within the vocal tract and in the near-field has also been visualized.

It has to be noted that the presented approach is not free of problems. For instance, imposing a frequency dependent impedance at the vocal tract walls in the time domain is not straightforward. Some efficient steps towards this goal can be found, for example, in Nieuwenhof and Coyette (2001). Moreover, with the proposed formulation acoustic coupling between the vocal tract and the subglottal cavities and the vocal folds cannot be considered. A solution to this problem may consist in solving the whole problem of vowel production in a single domain that includes all these elements (subglottal cavities, vocal folds and vocal tract). However, this might require the resolution of a more complex problem such as the Navier-Stokes equations for a compressible flow and consider fluid-structure interaction between the vocal folds and the acoustic field. Some works in this direction have recently been done (see e.g., Kaltenbacher *et al.*, 2014).

Chapter 3

Validation of numerical simulations against experiments

The aim of this chapter is to validate the FEM approach presented in Chapter 2 against experiments. In addition, the onset and behavior of high order propagation modes will be examined and the plane wave propagation regime will be evaluated. For that purpose, simplified vocal tract geometries with an increasing degree of complexity will be generated. Different magnitudes will be computed using these geometries, which include vocal tract transfer functions, pressure-pressure transfer functions and pressure distribution maps of a given formant or antiresonance. The results herein exposed will show that the proposed numerical approach can be used with confidence for vowel vocal tract acoustics, and will highlight the importance of three-dimensional acoustic modelling.

This chapter is partially based on the following works

- Rémi Blandin, Marc Arnela, Oriol Guasch, Rafael Laboissière, Xavier Pelorson, Annemie Van Hirtum and Xavier Laval, "Effects of higher order propagation modes in vocal tract like geometries," Journal of the Acoustical Society of America, Submitted.
- Rémi Blandin, Xavier Pelorson, Annemie Van Hirtum, Rafael Laboissière, Oriol Guasch and Marc Arnela (2014), "Effet des modes de propagation non plan dans les guides d'ondes à section variable," French Acoustical Conference 2014, April 22–25, Poitiers, France, pp. 745–751.
- Xavier Pelorson, Annemie Van Hirtum, Boris Mondet, Oriol Guasch and Marc Arnela (2013); "Three-dimensional vocal tract acoustics," Acoustics 2013, November 10–15, New Delhi, India.

3.1 Introduction

In this chapter the finite element strategy for vowel production presented in Chapter 2 will be validated. To do so, numerical simulations will be performed with a set of simplified vocal tract geometries for vowel /a/ and contrasted to experimental results. These geometries have been designed with an increasing degree of complexity. The most simple configuration consisting of two connected circular tubes with different cross-sections will be first considered. Second, simplified vocal tract geometries will be built from the area functions in Story (2008).

In order to examine the onset and behavior of higher order modes, the shape of these geometries will be generated as follows. First, circular vocal tracts with a centric configuration to connect each one of the cross-sections will be considered. This is of special interest since this configuration is widely used as a first approximation to generate three-dimensional geometries from area functions (see e.g., Vampola *et al.*, 2008a; Arnela and Guasch, 2014b). One can move forward and also consider eccentric junctions. This configuration should be slightly closer to the reality, since it better mimics the distribution of cross-sections along the vocal tract. The upper line that shares all cross-sections could be interpreted as if it was the palate. Finally, an additional case will be studied for the simplified vocal tracts. Elliptical cross-sections which can better approximate the vocal tract shape will be considered.

Numerical simulations and experiments will be performed for each configuration, from which different magnitudes will be extracted and compared. First, vocal tract transfer functions will be only computed for numerical simulations. Second, pressure-pressure transfer functions between different points within the vocal tracts will be simulated and contrasted to experimental results. Third, pressure maps will be calculated to examine the acoustic pressure distribution within the vocal tract and in the outer near-field for a given formant or antiresonance.

This chapter is organized as follows. In Section 3.2 the used methodology is presented. The generated vocal tract models are introduced in Section 3.2.1, the performance of numerical simulations and experiments is respectively specified in Section 3.2.2 and Section 3.2.3, and details on the computation of vocal tract transfer functions, pressure-pressure transfer functions and pressure maps are provided in Section 3.2.4. The obtained results are then examined in Section 3.3 for the two-tube configuration (Section 3.3.1) and the simplified vocal tracts (Section 3.3.2). Conclusions finally close the chapter in Section 3.4.

3.2 Methodology

3.2.1 Vocal tract models

We will first start our analysis with a very simplified model for the vocal tract of vowel /a/. This consists on a system made of two connected tubes with circular cross-sections. The cross-sectional area of the duct ending at the mouth exit is larger than the one having its input at the glottal cross-section (see e.g. Fant, 1970). The first tube is 85 mm in length and has an internal diameter of 14 mm, and the second one has the same length but with an internal diameter of 29.5 mm (see Fig. 3.1). Two different configurations are considered to connect them. In the first one the two tubes share the same central axis, while in the second one the junctions are eccentric and share a line through the upper cross-sections. Hereafter we will refer to these two configurations as the centric and eccentric cases, respectively.

Next we will increase the complexity of the geometries and consider simplified vocal tract geometries generated from the area functions in Story (2008). Three configurations are built for vowel /a/ (see Fig. 3.2). In the first one the vocal tract is generated with circular cross-sections and centric junctions. This case is of particular interest since it is widely used in the literature due to its simplicity (see e.g., Speed *et al.*, 2013; Arnela and Guasch, 2014b). The second case also uses circular cross-sections but they are connected

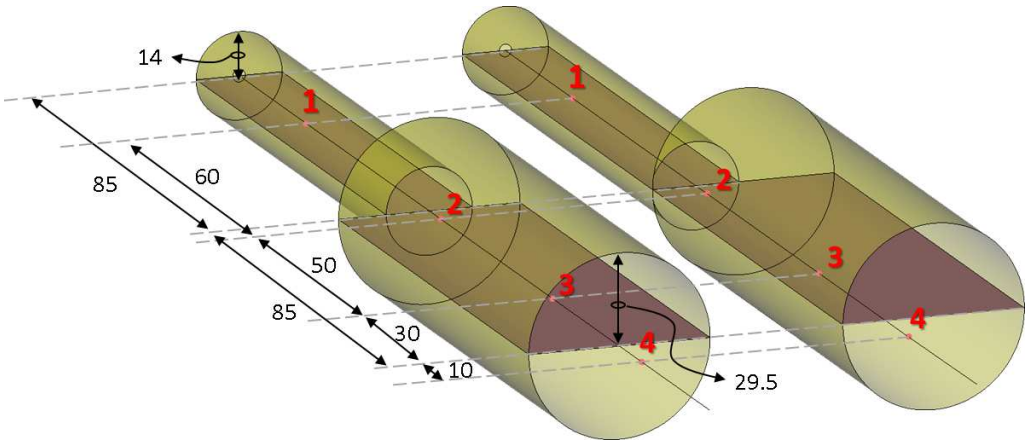


Fig. 3.1: Sketch with the tube system dimensions and with the points used to compute the pressure-pressure transfer functions $H_{ij}(f)$. Centric configuration (left) and eccentric configuration (right). In the inner part of the vocal tracts the midsagittal plane is highlighted. Units are expressed in millimeters.

with eccentric junctions, as done in the two-tubes configuration. Finally, in the third case an eccentric configuration is maintained but replacing the circular cross-sections with elliptical shapes. The eccentricity of the mouth for vowel /a/, obtained from Fromkin (1964), is applied to all cross-sections to generate an elliptical shape (as done in Arnela *et al.*, 2013). From now on these three configurations will be respectively termed for brevity as the circular-centric, circular-eccentric and elliptical-eccentric cases. Although the three configurations are rough approximations of a realistic vocal tract, they can be interpreted as approximations of increasing degree of reality. The circular-centric case is the most simple configuration. With eccentric junctions we obtain a common upper-line between the cross-sections that can be viewed as corresponding to the hard palate (circular-eccentric case), and including elliptical cross-sections (elliptical-eccentric case) we can obtain a closer vocal tract shape than with circles.

In order to generate the vocal tract geometries with eccentric junctions (circular-eccentric and elliptical-eccentric cases), one can simply start from a centric configuration and move all cross-sections up to make them share an upper-line. However, by using this methodology a longer vocal tract midline (the line or curve that pass though the center of each cross-section) is obtained, which can produce a shift down of the formant locations. So to compensate this effect, the vocal tract lengths are corrected so that all configurations have the same midline length than the centric case.

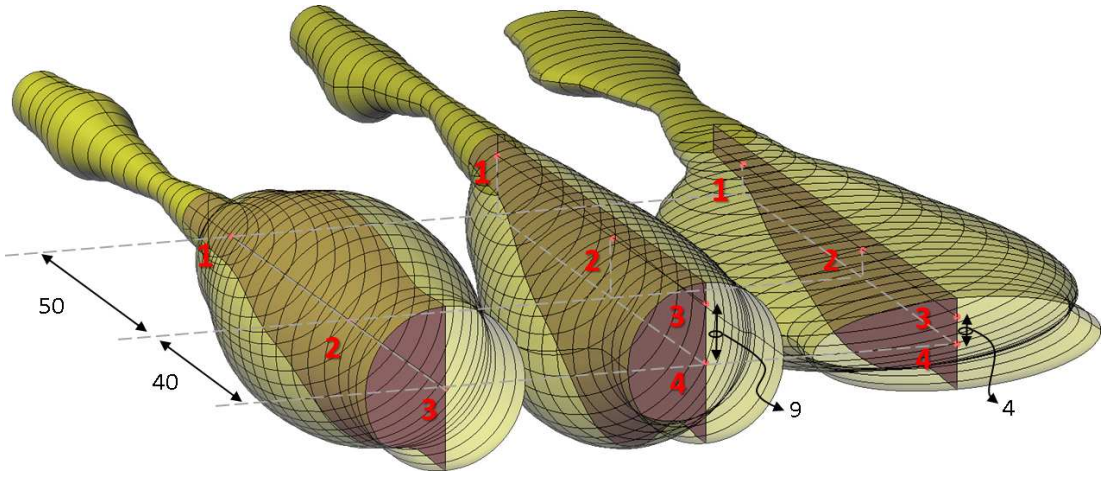


Fig. 3.2: The three simplifications of the vocal tract for vowel /a/ including the points used to compute the pressure-pressure transfer functions $H_{ij}(f)$. (left) circular-centric, (center) circular-eccentric and (right) elliptical-eccentric cases. In the inner part of the vocal tracts the midsagittal plane is highlighted. Units are expressed in millimeters.

3.2.2 Finite Element simulations

The Finite Element Method (FEM) is used to solve the time domain wave equation for the acoustic pressure (see Chapter 2). First, each one of the vocal tract models presented in Section 3.2.1 are set in a rigid flat baffle with dimensions 0.3 m x 0.3 m. This baffle constitutes one side of a rectangular volume of 0.3 m x 0.3 m x 0.2 m in size which allows sound waves radiate out from the vocal tract. Free-field radiation conditions are then emulated by means of a Perfectly Matcher Layer (PML) of width 0.1 m that is surrounding this radiation space and absorbs the outgoing sound waves. Details about the PML configuration and the mesh description can be found in Section 2.4.

In what concerns boundary conditions, a Gaussian pulse is imposed at the vocal tract entrance to compute transfer functions while a sinusoid with the frequency of interest is used for obtaining a pressure map. At the vocal tract walls, a constant frequency boundary admittance coefficient of $\mu = 0.0025$ is imposed to introduce losses. This value has been deemed appropriate after a tuning process where the resonance bandwidths obtained from simulations were adjusted to fit those from experiments. The baffle where the vocal tracts are set is considered rigid, i.e. $\mu = 0$ on its surface.

A FEM simulation lasting 25 ms for each case is then performed with a sampling rate of $f_s = 1/\Delta t = 2000$ kHz. A value of $c_0 = 344$ m/s is chosen for the speed of sound which corresponds to the average temperature value of 21°C reported during experiments.

3.2.3 Experiment description

In what concerns the experimental setup (see Fig. 3.3, extracted from Blandin *et al.*, 2014a), a mechanical replica made of Plexiglas is used for the two-tubes while the three simplified vocal tracts are built with a 3D printer (ProJet 3510 SD). The exit cross-section of these replicas are set on a rigid plane baffle of dimensions 365 mm x 360 mm. The resulting models are then located inside a sound-insulated room (1.92 m x 1.95 m x 1.99 m, Volume = 7.45 m^3, Van Hirtum and Fujiso, 2012).

The compression chambers Monacor KU-916T and Eminence PSD:2002S-8 are respectively used for the frequency ranges $100 - 2000$ Hz and $2000 - 10000$ Hz as a sound source. Both compression chambers are located outside the sound-insulated room to avoid non-desired interferences. They are connected with the entrance of the replica by means of an adaptation device, which radiates the sound inside the replica through a 2 mm diameter hole (this hole is also reproduced in the FEM simulations).

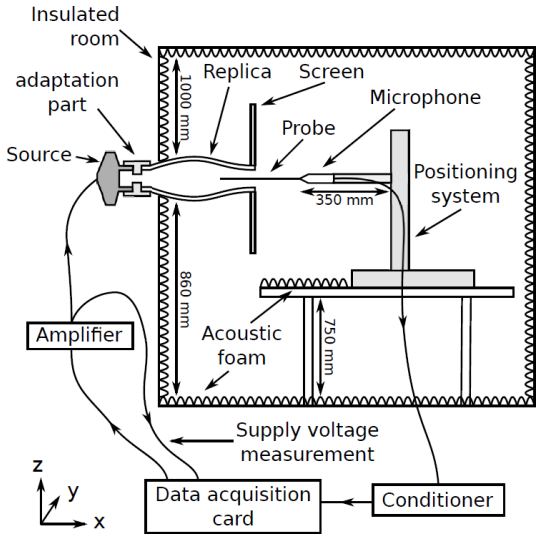


Fig. 3.3: Experimental setup (courtesy of Rémi Blandin and Xavier Pelorson).

To measure the acoustic pressure a sound probe (Brüel and Kjaer 4182) is used, which can be located inside and outside the replicas using a 3D stage positioning system (OWIS PS35) with an accuracy of ± 4 μm. The initial position of this probe is however less accurate (1 mm) and can be an error source which produces significant differences between simulations and experiments. The positioning system is set on a table inside the insulated room, which is covered with acoustic foam to minimize undesired sound reflections.

All measurements were performed in the GIPSA-lab facilities in Grenoble, France, by Rémi Blandin, Xavier Pelorson, Annemie Van Hirtum and Xavier Laval.

3.2.4 Computation of transfer functions and pressure maps

First, vocal tract transfer functions are computed from FEM simulations to study the acoustic behavior of each vocal tract model. A Gaussian pulse is used to generate the volume velocity $Q_g(t)$ imposed at its entrance. Collecting the acoustic pressure $p_o(t)$ at the mouth exit, the vocal tract transfer function is computed as

$$H(f) = P_o(f)/Q_g(f), \tag{3.1}$$

with $P_o(f)$ and $Q_g(f)$ standing for the Fourier Transform of the acoustic pressure $p_o(t)$ and the volume velocity $Q_g(t)$, respectively.

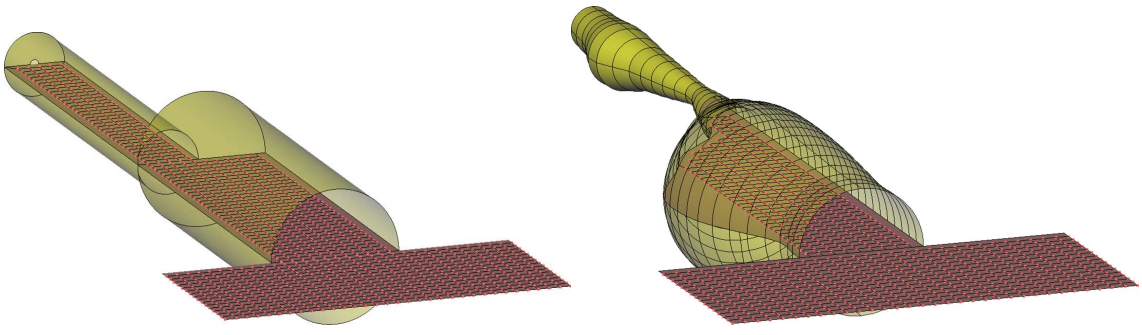
Second, pressure-pressure transfer function between two arbitrary points i and j are

obtained for both FEM and experiments. This can be computed as

$$H_{ij}(f) = P_j(f)/P_i(f), \tag{3.2}$$

where $P_i(f)$ and $P_j(f)$ are the Fourier Transform of the acoustic pressure at points i and j, respectively. The selected points can be observed in Fig. 3.1 for the two-tubes and in Fig. 3.2 for the simplified vocal tracts. This magnitude is of special interest for the realization of experiments, since the microphone and source calibration can be avoided; a single microphone can be used and the influence of the experimental setup (amplifier, wires and microphone conditioner) in the measure can be neglected.

Finally, to illustrate the importance of high-order modes and to compare simulations against experiments, the pressure field at a given frequency within the vocal tract and in the near-field is also computed. To do so, the vocal tract is excited at the glottis cross-section with a sinusoidal signal. Its frequency is chosen to coincide with a vocal tract resonance or antiresonance, which can be obtained from the spectrum of the acoustic pressure at the vocal tract exit after exciting its entrance with a broadband signal. Since small differences can be produced between experiments and simulations, these values are extracted for each method to ensure similar pressure distributions. The acoustic pressure is then tracked in a structured grid of points with a distance of 2.5 mm apart located in the midsagittal plane. An example of this grid for the eccentric two-tubes and the circular-eccentric /a/ can be observed in Fig. 3.4. The points have been selected so that the microphone probe can reach any position (the probe can not be bent). Note than in contrast to the two-tubes case (Fig. 3.4a), the grid of points for the circular-eccentric /a/ (Fig. 3.4a) can only cover a small region of the midsagittal cut (shadowed region in Fig. 3.4). The mean absolute value of the acoustic pressure is finally computed. In the case of simulations, this is done by using the values from the last 5 ms.



(a) eccentric two-tubes (b) circular-centric /a/

Fig. 3.4: Grid of points (in red) where the acoustic pressure is extracted to compute the formant pressure patterns. The shadowed region corresponds to the midsagittal cut.

3.3 Results

3.3.1 Two-tubes

Vocal tract transfer functions

First, vocal tract transfer functions are computed in simulations for the centric and eccentric cases, collecting the output acoustic pressure $p_o(t)$ at the exit of the larger tube. In Fig. 3.5 the obtained vocal tract transfer functions are presented. First, it can be observed that by using an eccentric configuration the pressure levels are increased with an offset about 15dB. Second, it can be appreciated that for the centric case, higher order modes are not visible. This is due to the symmetry of the geometry. In contrast, if we break this symmetry and consider an eccentric configuration, higher order modes become apparent for frequencies beyond 5 kHz. Resonances with asymmetric bandwidths and antiresonances can be observed in this frequency range. Below 5 kHz there are not significant differences between the centric and eccentric cases. Therefore it can be said that for this particular case the plane wave assumption holds up to about 5 kHz.

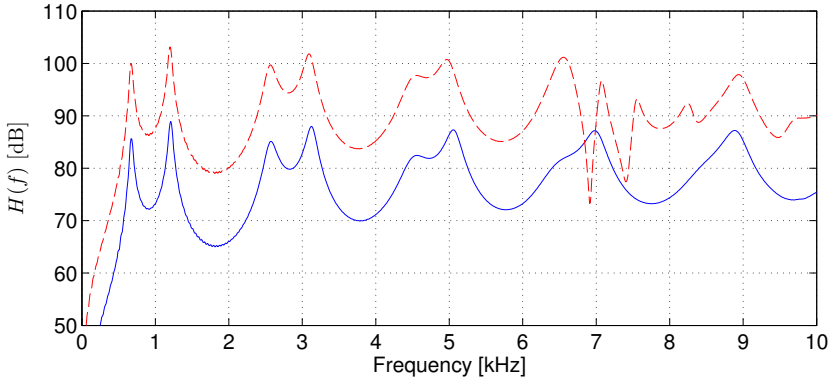
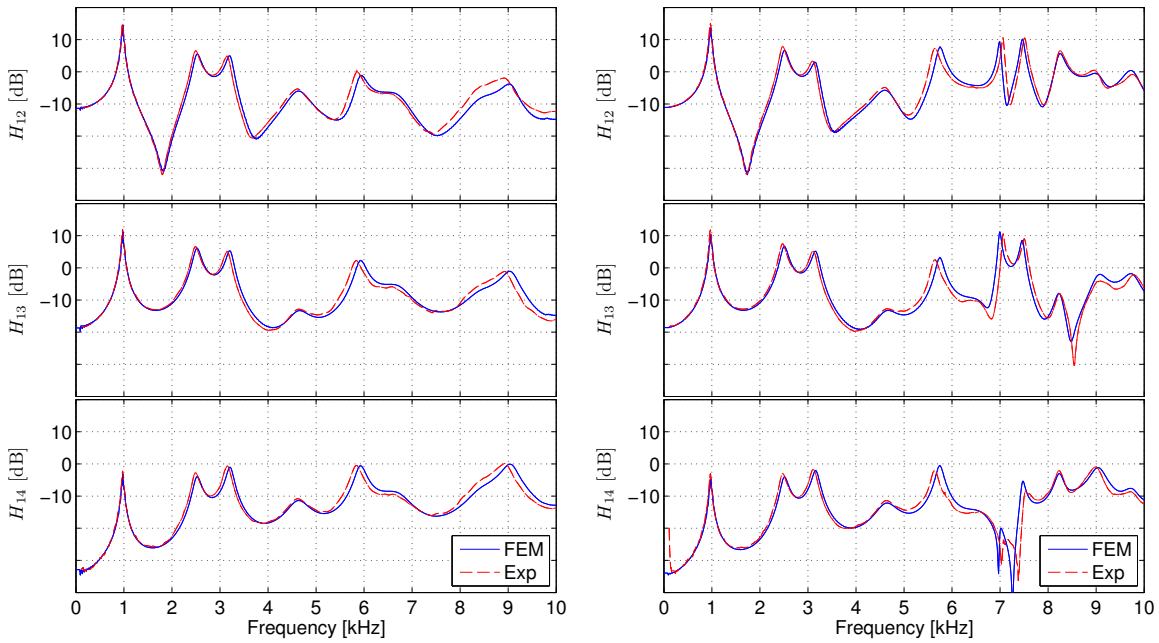


Fig. 3.5: Vocal tract transfer functions computed in the FEM numerical simulations of the two-tubes. (— Centric case, - - Eccentric case)

Pressure-pressure transfer functions

For both, simulations and experiments, pressure-pressure transfer functions are next calculated from a set of four points distributed along the axis of the smaller tube (see Fig. 3.1). In Fig. 3.6 pressure-pressure transfer functions for the centric and eccentric cases are shown. The figures also contain the results obtained from experiments for comparison.



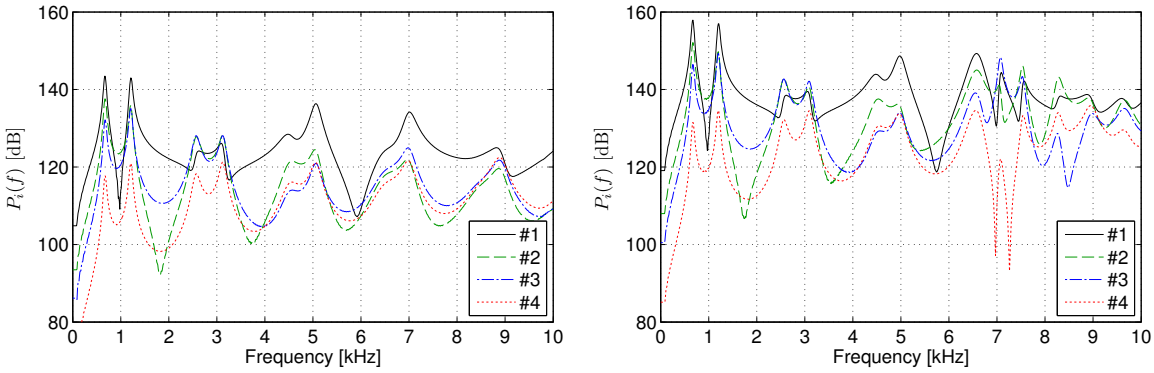
(a) Centric two-tubes (b) Eccentric two-tubes

Fig. 3.6: Modulus of the pressure-pressure transfer functions $H_{ij}(f)$ for the two-tubes obtained in Finite Element simulations (FEM) and experiments (Exp).

All transfer functions are computed taking point #1, located within the narrowest tube, as the reference. Then, the transfer functions H_{12}, H_{13} and H_{14} are respectively computed using point #2 located in the largest tube and close to the junction, point #3 close to the mouth exit but still inside the larger tube, and point #4 which is outside the vocal tract, in the near-field.

The first observation that can be made is that for any of the pressure-pressure transfer functions both simulations and experiments match in the examined frequency range (0,10) kHz, including the high frequency region where non-planar waves can also propagate. This validates the proposed FEM approach for this case.

Let us now focus in the results for the centric configuration (Fig. 3.6a). Moving from point #2 (H_{12}) to point #3 (H_{13}), it can be appreciated that in the frequency region below 5 kHz large differences are found between figures, while above this value no significant changes are appreciated. This behavior can be expected for planar waves, since for low frequencies the pressure field distribution in the propagation direction can contain abrupt amplitude changes while for high frequencies they will be smoother. Moving next from point #3 (H_{13}) to point #4 (H_{14}), at first sight no large differences are observed. However, looking carefully at the amplitude levels it can be appreciated that H_{14} is gradually



(a) Centric two-tubes (b) Eccentric two-tubes

Fig. 3.7: Pressure spectrum $H_i(f)$ for the two-tubes at the different points $i = 1, 2, 3, 4$ (see Fig. 3.1 and Fig. 3.2 for the point locations in the centric and eccentric cases respectively).

decreasing its energy towards low frequencies. See for instance that the pick at 1 kHz has $\sim$ 10 dB and 0 dB for H_{13} and H_{14} respectively, while the pick at 9 kHz has a similar amplitude for both magnitudes. This behavior can be attributed to a radiation effect, since the power radiated outwards is proportional to the radiation resistance which increases with frequency (see e.g., Kinsler *et al.*, 2000). Remember that point #3 is located within the largest tube close to its exit and point #4 is outside.

In what concerns the eccentric case (see Fig. 3.6b), it behaves similar to the centric configuration in the low frequency region. However, in this case for frequencies above 5 kHz significant changes occur between H_{12}, H_{13} and H_{14}. Complex pressure fields may be given within the tube system, explaining the difference between H_{12} and H_{13}, while the variation between H_{13} and H_{14} can be attributed to a complex directivity pattern. Therefore, the evidence of higher order modes beyond 5 kHz is again confirmed for the eccentric case, as observed before in the vocal tract transfer functions (see Fig. 3.5).

Finally, we would like to note that one does not have to confuse the picks in the pressure-pressure transfer functions as vocal tract resonances or formants, appearing for instance in the vocal tract transfer functions shown in Fig. 3.5 (see e.g., Guasch and Magrans, 2004; Guasch, 2009, for an explanation of peaks and dips in mechanical transmisibiliy functions). Looking at the pressure spectrums of each one of the four points (see Fig. 3.7), note for instance that the pick at $\sim$ 1 kHz in the pressure-pressure transfer functions (see Fig. 3.6) is produced by the minimum pressure value of point #1 at this frequency (see Fig. 3.7a or Fig. 3.7b) when $H_{1j} = H_j/H_1$ is computed, with $j = 2,\ 3$. Moreover, also observe that the first two typical resonances of the vowel /a/ are canceled since they appear in all four points (see Fig. 3.7).

Pressure maps

Finally, in order to illustrate the three-dimensional behavior of acoustic waves, the pressure pattern of a given formant has been obtained, within the duct system and in the near-field, in a structured grid of points. One formant in the low frequency region has been chosen, where the plane wave assumption holds, and another one beyond this limit. The selected formants are the third formant (F3) and the eight formant (F8), for both the centric and eccentric configurations. In Table 3.1 their frequencies are presented. These values have been extracted for simulations from the vocal tract transfer functions in Fig. 3.5, while for experiments they have been computed from a frequency sweep of the duct system over the analyzed frequency range. Note from Table 3.1 that there are small variations between FEM and experiments. However, these deviations are less than $3-5\%$, so they can be considered perceptually irrelevant (Flanagan, 2008).

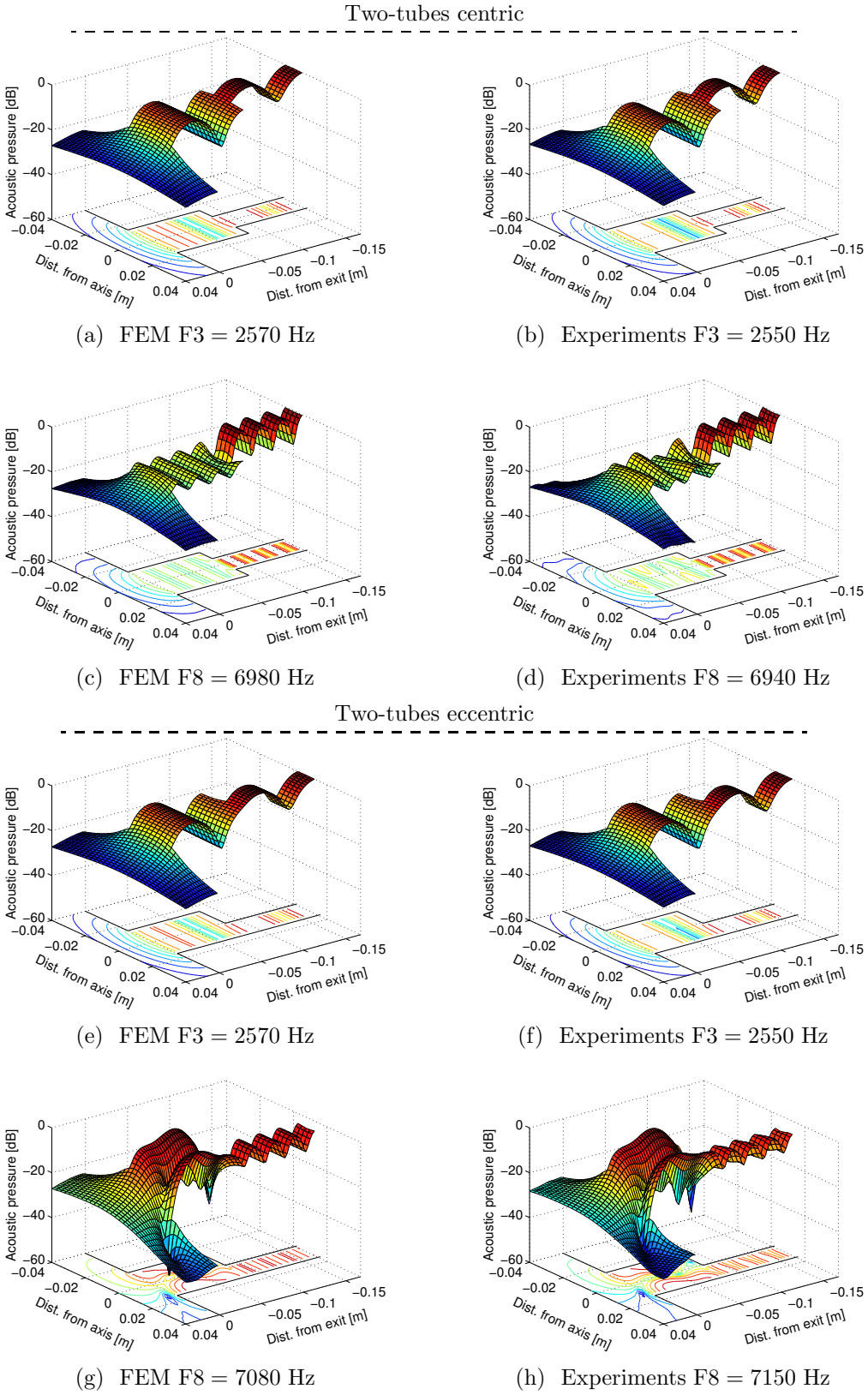
Figure 3.8 presents the formant patterns of F3 and F8 for both the centric and eccentric configurations obtained in FEM and experiments (Exp). The origin of coordinates has been defined at the tube exit. Negative values in the x-axis (longitudinal or propagation direction) will then correspond to the interior of the tube and positive ones to the outer near-field. The y-axis represents the transverse direction, with the zero value denoting the center of the larger tube. The z-axis denotes mean acoustic pressure amplitude in dB, normalized to the maximum pressure level. Moreover, to facilitate comparison between FEM and experiments, two additional graphics are represented in Fig. 3.9 for each one of the formants. They correspond to cuts of the pressure maps shown in Fig. 3.8 in the sagittal plane (propagation direction) and just in front of the mouth exit in the frontal, or coronal, plane (transverse direction).

Table 3.1: Locations of the third formant (F3) and the eight formant (F8) obtained from FEM simulations and experiments (FEM/Experiments). Values are expressed in Hz.

	F3	F8
centric	2570 / 2550	6980 / 6940
eccentric	2570 / 2550	7080 / 7150

Let us first focus on the results for the centric case. If we look at the pressure map for F3 obtained for both FEM and experiments (see Fig. 3.8a and Fig. 3.8b), it can be observed that a similar acoustic pressure pattern is obtained for both methods. Note that pressure variations only take place in the longitudinal direction within the duct system, and that the radiated field propagates as spherical waves. Analyzing the pressure map cuts for this frequency (see Fig. 3.9), similar results are obtained for FEM and experiments in the sagittal cut (see Fig. 3.9a), and also a similar spherical pattern is observed in the frontal cut (see Fig. 3.9b). This behavior corresponds to plane wave propagation. If we now move to F8, the obtained pressure patterns for FEM and experimental results are slightly different (see Fig. 3.8c and Fig. 3.8d). In FEM simulations plane waves are mainly produced, while in experiments a higher order mode is also excited. This is attributed to the fact that in experiments the geometry is not perfectly symmetric, which produces the onset of a higher order mode. If we analyst the acoustic pressure in the sagittal cut (Fig. 3.9c), we set very close results between FEM and experiments, since this higher order mode has a nodal line in the tube center line and therefore, only the planar wave becomes visible. Comparing with the same cut at F3 (see Fig. 3.9a), as one could expect the amplitude variation within the duct system is smaller for F8, since for high frequencies planar waves have smoother pressure distributions than in the lower frequency region, as mentioned before during the pressure-pressure transfer function analysis. In what concerns the radiation pattern, spherical wave propagation is also observed for both experiments and FEM (see Fig. 3.9d). Compared to F3 (see Fig. 3.9b), the radiation pattern is slightly more directive. The main lobe at -0.04 m reaches -24 dB for F3 while for F8 it has a value of -26 dB (remember that the pressure levels are normalized to the maximum value of the pressure maps).

With regard to the eccentric configuration, a similar behavior to the centric case is observed for F3. Compare the pressure maps for the eccentric case (Fig. 3.8e and Fig. 3.8f) with the centric case (Fig. 3.8a and Fig. 3.8b), and their respective cuts (Fig. 3.9e and Fig. 3.9f with Fig. 3.9a and Fig. 3.9b). For this frequency we are still under plane wave propagation, so the duct configuration does not have any influence. Plane waves propagate within the duct whereas spherical propagation is generated at the exit. These results contrast with those obtained for F8 (see pressure maps in Fig. 3.8g and Fig. 3.8h). In this case a more complex pressure pattern is given within the duct system, and there is no longer spherical radiation. It becomes then apparent the presence of higher order modes for this frequency regime. If we look now at the cuts (see Fig. 3.9g and Fig. 3.9h), although FEM and experiments do not match as good as in previous comparisons, close results are still obtained between them.



(a) FEM F3 = 2570 Hz

(b) Experiments F3 = 2550 Hz

(c) FEM F8 = 6980 Hz

(d) Experiments F8 = 6940 Hz

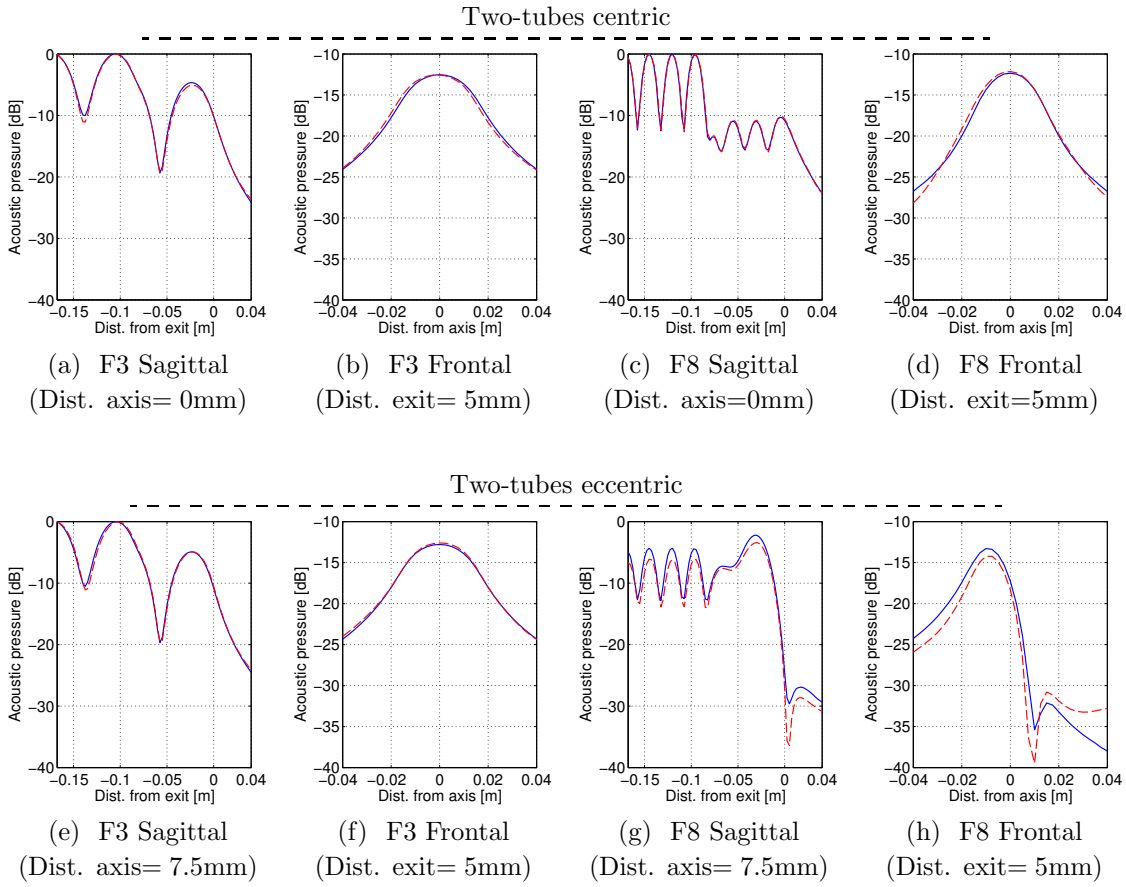
(e) FEM F3 = 2570 Hz

(f) Experiments F3 = 2550 Hz

(g) FEM F8 = 7080 Hz

(h) Experiments F8 = 7150 Hz

Fig. 3.8: Pressure maps within the duct system with centric and eccentric configuration for the third formant (F3) and the eight formant (F8).



(a) F3 Sagittal (Dist. axis= 0mm) (b) F3 Frontal (Dist. exit= 5mm) (c) F8 Sagittal (Dist. axis=0mm) (d) F8 Frontal (Dist. exit=5mm)

(e) F3 Sagittal (Dist. axis= 7.5mm) (f) F3 Frontal (Dist. exit= 5mm) (g) F8 Sagittal (Dist. axis= 7.5mm) (h) F8 Frontal (Dist. exit= 5mm)

Fig. 3.9: Cuts of the pressure maps (see Fig. 3.8) in the sagittal and frontal planes of the duct system with centric and eccentric configuration for the third formant (F3) and the eight formant (F8).

3.3.2 Simplified vocal tracts

Vocal tract transfer functions

In Fig. 3.10 the computed vocal tract transfer functions for the three configurations in Fig. 3.2 are shown. It can be appreciated that the use of an eccentric configuration increases the pressure level by ~ 20 dB compared to the centric configuration, similar to what was observed for the two-tubes (see Section 3.3.1). On the other hand, below 5 kHz no significant differences are produced between cases in terms of formant location and bandwidth. Therefore, it can be asserted that below this value the examined vocal tract is under plane-wave propagation regime, since the introduced geometry modifications in the vocal tract shape are not affecting its behavior. Note also that the formant locations in this regime match thanks to the applied vocal tract length correction in the eccentric configurations. Beyond 5 kHz, higher order modes are excited and large differences take plane between cases, specially when moving from a centric to an eccentric configuration. See for instance that with an eccentric configuration antiresonances appear, while they are missing for the centric case due to the radial symmetry. However, comparing the elliptical-eccentric with the circular-eccentric not so large differences can be appreciated. The resonances and antiresonances are moved to higher frequencies and larger bandwidths are obtained when elliptical shapes are considered.

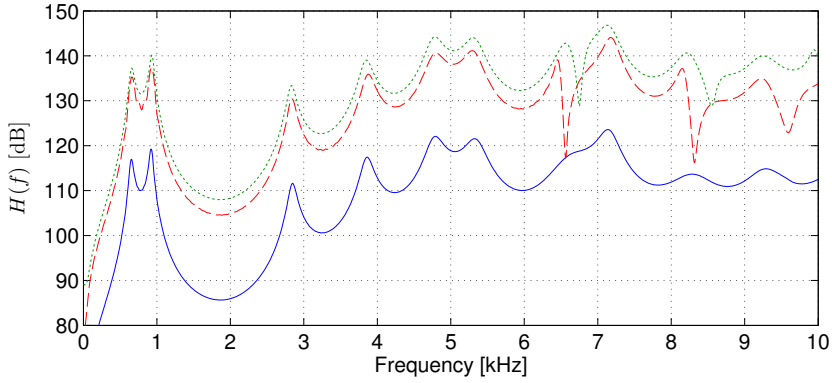


Fig. 3.10: Vocal tract transfer functions computed in the FEM numerical simulations of the simplified vocal tracts. (— Circular-centric, - - Circular-eccentric, ··· Elliptical-eccentric)

Pressure-pressure transfer functions

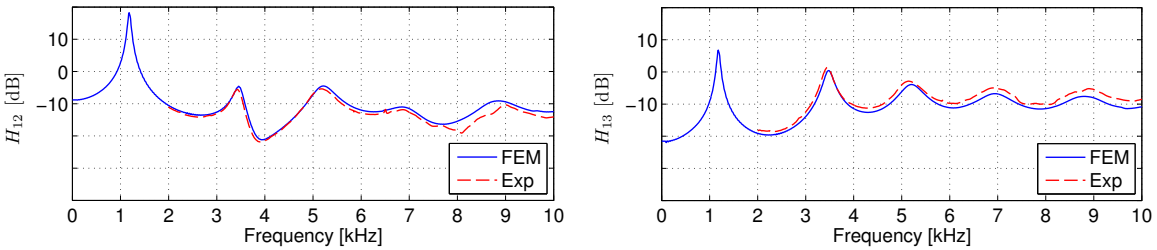
Pressure-pressure transfer functions are computed in a set of three points located within the vocal tract and distributed along the longitudinal axis centered with the glottal cross-

section (see Fig. 3.2). The first point #1 is located within the narrow part of the vocal tract, point #2 is in the widest region, and point #3 is placed in the mouth cross-sectional area. For the eccentric configurations an additional point (point #4) is also considered to capture the pressure field in the center of the mouth exit. Note that this point is the equivalent to point #3 for the centric configuration. The obtained transfer functions are shown in Fig. 3.11.

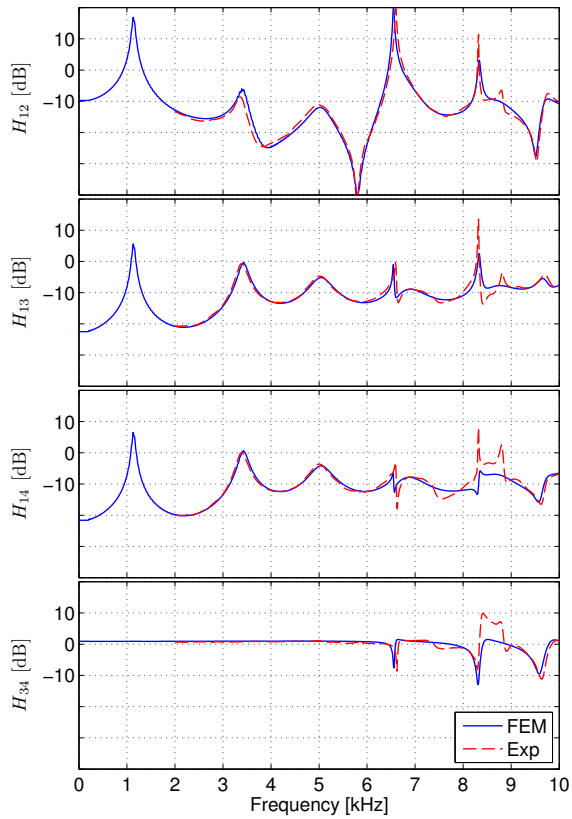
In what concerns the circular-centric case (see Fig. 3.11a), comparing H_{12} with H_{13} it can be appreciated that the most significant differences are produced below 5 kHz, as already observed for the two-tubes (see Section 3.3.1). This can be attributed again to the typical behavior of planar waves, which in terms of amplitude have smoother pressure field distributions at high frequencies than in the lower ones, the point location making less sensitive in the high frequency range.

With regard to the circular-eccentric configuration (see Fig. 3.11b), a similar behavior than in the circular-centric case is observed, for the low frequency region, when comparing H_{12} with H_{13}. However, in this case the higher order modes become apparent beyond 5 kHz. If we now compare H_{13} with H_{14} (an equivalent transfer function that considers the point #4 centered in the mouth exit instead of the non-centered point #3) it can be observed that in the low frequency region a similar transfer function is obtained. However, above 5 kHz significant differences are produced. See for instance that the resonance at about 8 kHz has changed its shape and that the resonance in H_{13} at about 9.5 kHz is now an antiresonance in H_{14}. This illustrates again the effect of higher order modes, which have complex pressure distribution fields not only in the propagation direction but also in the radial one. This behavior can be easily observed computing the transfer function between these two points, H_{34}. Below 5 kHz a flat response is obtained, indicating that the acoustic pressure is the same in points #3 and #4. However, beyond this value the pressure field changes in the radial direction, which justify the appearance of the antiresonances observed in the figure.

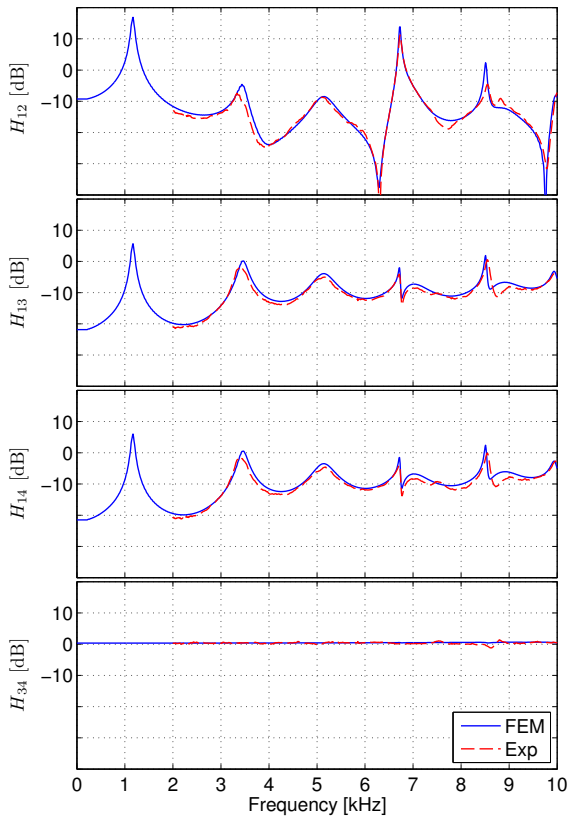
The elliptical-eccentric case (see Fig. 3.11c) behaves similar to the circular-eccentric configuration. However, in this case no significant differences can be appreciated between H_{13} and H_{14}. This is also observed in the flat response that H_{34} has in the whole examined frequency range, indicating that the pressure distribution is not changing in this direction, in contrast to the circular-eccentric case. This is a result that could be expected since for an elliptical configuration the midsagittal distance is smaller than for a circular one (see Fig. 3.2). However, similar results to the circular-eccentric case may be expected if we examine the major semi-axis instead of the minor semi-axis.



(a) Circular-centric



(b) Circular-eccentric



(c) Elliptical-eccentric

Fig. 3.11: Modulus of the pressure-pressure transfer functions $H_{ij}(f)$ for the simplified vocal tracts of vowel /a/ obtained in Finite Element simulations (FEM) and experiments (Exp).

On the other hand, in general both FEM simulations and experiments match for all the considered cases in the examined frequency range. Some discrepancies can be found in Fig. 3.11 beyond 8 kHz which are attributed to small measurement errors or imperfections of the vocal tract replicas.

Pressure maps

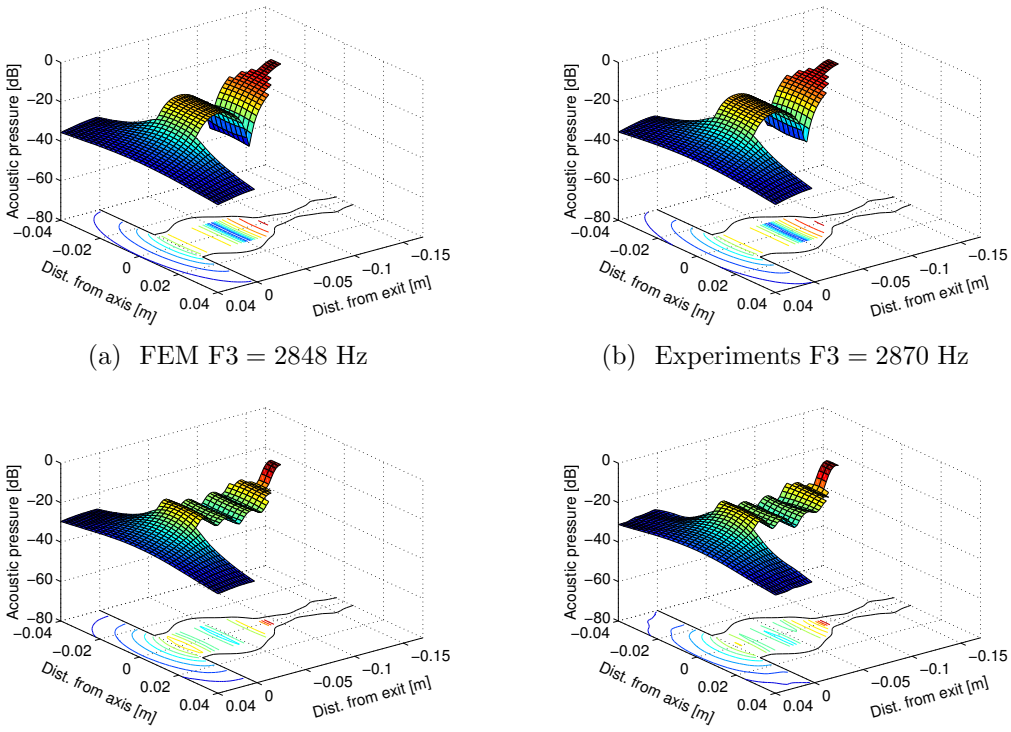
Finally, pressure maps are computed in a set of resonances and antiresonances belonging to each vocal tract configuration. Their locations are extracted in FEM from the vocal tract transfer functions in Fig. 3.10, whereas a frequency sweep over the analyzed frequency range has been performed in experiments to find these values. The selected frequencies are listed in Table 3.2. The third formant (F3) is chosen to analyze the low frequency range, while for high frequencies the eighth formant (F8) is used. The second antiresonance (A2) is also considered as an additional case for the circular-eccentric configuration.

Table 3.2: Formant (F_i) and antiresonance (A_i) frequency values used to compute pressure patterns in each vocal tract configuration for FEM simulations and experiments (FEM/Experiments). Values are expressed in Hz.

	F3	F8	A2
circular-centric	2848 / 2870	7139 / 7380	-
circular-eccentric	2841 / 2820	7178 / 7390	8315 / 8280
elliptical-eccentric	2833 / 2810	7130 / 7434	-

The obtained pressure maps are shown in Fig. 3.12, Fig. 3.13 and Fig. 3.14 for the circular-centric, circular-eccentric, and elliptical-eccentric cases, respectively. As done in the two-tubes (see Section 3.3.1), the pressure maps have been normalized to its maximum value. Some pressure map cuts in the sagittal plane and in the frontal plane just outside of the mouth exit are also shown in Fig. 3.15 to better compare FEM with experiments.

Focusing first in the pressure maps for the circular-centric configuration (see Fig. 3.12), a similar behavior than for the two-tubes is observed (see Section 3.3.1). For both, the low frequency formant F3 (see Fig. 3.12a or Fig. 3.12b) and the high frequency formant F8 (see Fig. 3.12c or Fig. 3.12d), plane waves are generated within the vocal tract, which are radiating outside the vocal tract following a spherical pattern. Comparing experiments with simulations, for F3 no differences can be appreciated. However, it seems again that in addition to the planar wave a higher order mode is also slightly excited for F8 in experiments (see Fig. 3.12d), in contrast to simulations (see Fig. 3.12c). This can be attributed again to small imperfections of the vocal tract replica. Some more details on the comparison between simulations and experiments are given in the pressure map cuts shown in Fig. 3.15. It can be appreciated that both FEM and experiments match in the circular-centric case for both formants, showing also for F8 a smaller dynamic range (see Fig. 3.15c) and a slightly higher directional behavior of the wave radiation (see Fig. 3.15d), with respect to F3 (see respectively Fig. 3.15a and Fig. 3.15b).



(a) FEM F3 = 2848 Hz

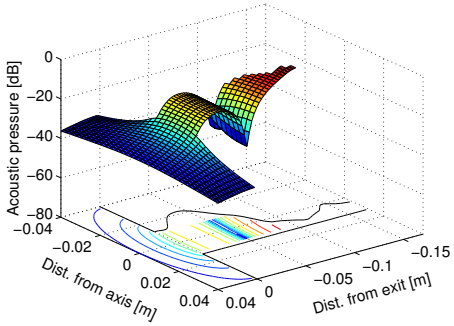
(b) Experiments F3 = 2870 Hz

(c) FEM F8= 7139 Hz

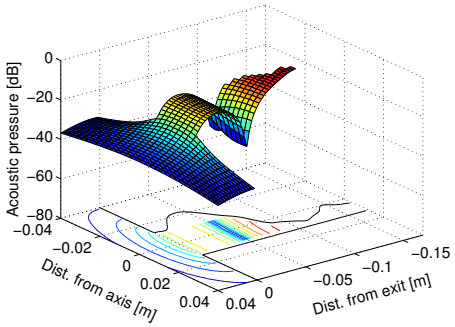
(d) Experiments F8= 7380 Hz

Fig. 3.12: Pressure maps in the circular-centric /a/ for the third formant (F3) and the eight formant (F8).

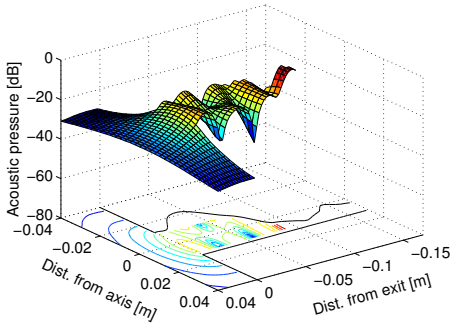
With regard to the circular-eccentric configuration, the pressure distribution for the low frequency formant F3 (see Fig. 3.13a or Fig. 3.13b) behaves similar to the circular-centric case. However, for F8 a complex pattern is produced within the duct due to the appearance of a higher order mode. In this case, the radiation pattern is still spherical, in contrast to what was observed for the high frequency formant of the two-tube configuration with eccentric junctions (see Fig. 3.8g). The opposite occurs for the second antiresonance A2 (see Fig. 3.13e or Fig. 3.13f). In this case the radiated pressure appears as strongly directional, in addition to the complex pressure field distribution. Comparing simulations with experiments, for both F3 and F8 no significant differences are appreciated. However, the radiation pattern for A2 is slightly different, although the pressure field distribution seems similar within the vocal tract. If we look at the pressure map cuts (see Fig. 3.15), it is confirmed that both simulations and experiments match for F3 (see Fig. 3.15e and Fig. 3.15f) and F8 (see Fig. 3.15g and Fig. 3.15h), and that A2 has a similar pressure distribution within the vocal tract (see Fig. 3.15i) although the radiation pattern is different (see Fig. 3.15j). While for simulations two non-symmetrical lobes are obtained, the measurements present an almost symmetrical behavior.



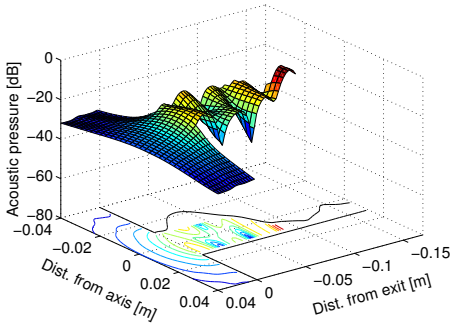
(a) FEM F3 = 2841 Hz



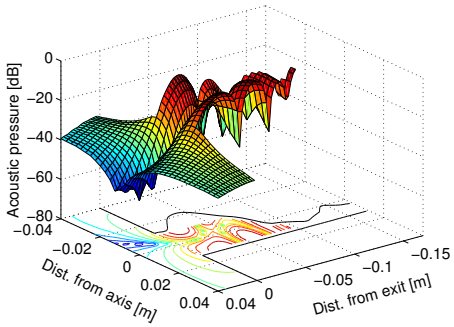
(b) Experiments F3 = 2820 Hz



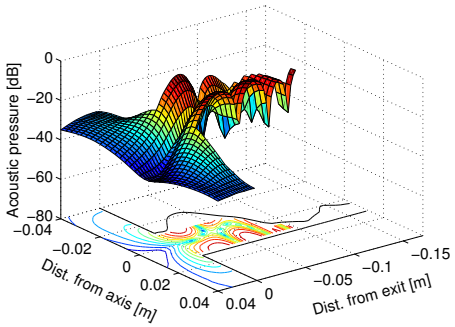
(c) FEM F8 = 7178 Hz



(d) Experiments F8 = 7390 Hz



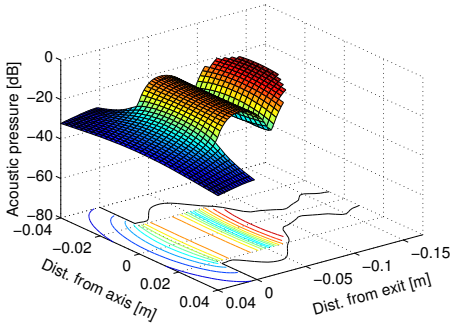
(e) FEM A2 = 8315 Hz



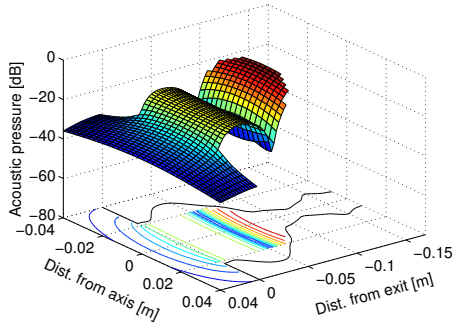
(f) Experiments A2 = 8280 Hz

Fig. 3.13: Pressure maps in the circular-eccentric /a/ for the third formant (F3), the eight formant (F8) and the second antiresonance (A2).

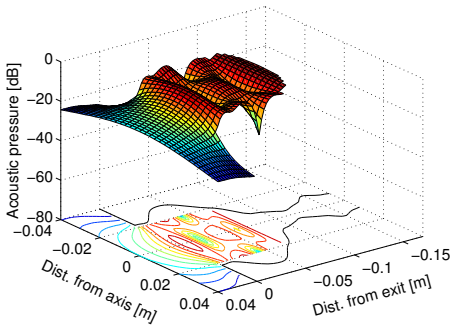
Finally in Fig. 3.14 the pressure maps for the elliptical-eccentric configuration are shown. It should be mentioned that in this case the pressure maps have been computed in the transverse plane (direction of the major semi-axis) that contains the mouth exit center. This plane is perpendicular to the midsagittal plane where the pressure maps for the circular-centric and circular-eccentric cases have been obtained. In what concerns the low frequency formant F3 (see Fig. 3.14a or Fig. 3.14b), plane waves are again appreciated along the vocal tract which spherically radiates outwards. In contrast, for the high frequency formant F8 (see Fig. 3.14c or Fig. 3.14d) a complex pressure distribution is generated within the vocal tract, although spherical waves are still emanating from its mouth aperture. Both simulated and experimental pressure maps show similar results. The same conclusion can be derived when analyzing the pressure map cuts for F3 (see Fig. 3.15k and Fig. 3.15l) and F8 (see Fig. 3.15m and Fig. 3.15n), although a small level offset is observed for F3. This is probably a consequence of the amplitude scaling performed for all pressure maps, which were normalized to the maximum pressure level to allow a fair comparison between FEM and experiments.



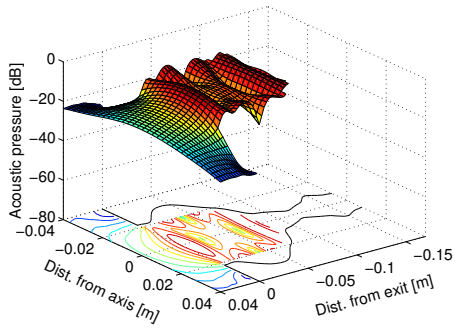
(a) FEM F3 = 2833 Hz



(b) Experiments F3 = 2810 Hz



(c) FEM F8 = 7130 Hz



(d) Experiments F8 = 7434 Hz

Fig. 3.14: Pressure maps in the elliptical-eccentric /a/ for the third formant (F3) and the eight formant (F8).

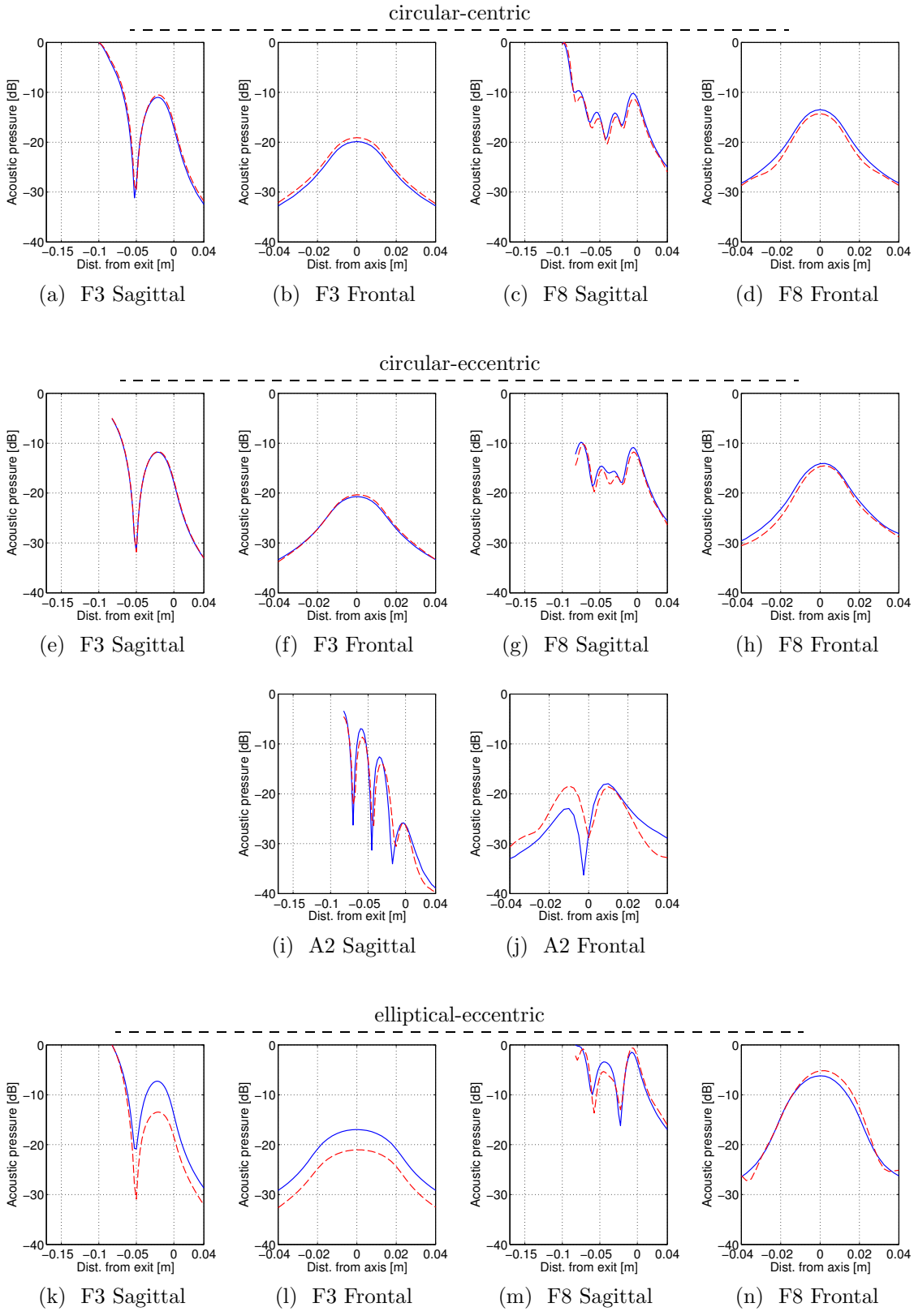


Fig. 3.15: Cuts of the pressure maps in the sagittal plane (Dist. from axis = 0 mm) and in the frontal plane (Dist. from exit = 5 mm) for the three vocal tract configurations.

3.4 Conclusions

In this chapter the FEM approach presented in Chapter 2 has been validated against experimental results. Simplified vocal tract geometries for the production of vowel /a/ have been studied for this purpose, starting from the most simple two-tubes configuration and moving to vocal tracts constructed from area functions available in the literature. These vocal tracts have been generated with different configurations, including centric and eccentric junctions and considering circular and elliptical shapes.

Vocal tract transfer functions have been first simulated, showing that higher order modes can be produced beyond 5 kHz for those configurations that have eccentric junctions and circular or elliptical cross-sections. Next, pressure-pressure transfer functions have been obtained for both FEM and experiments. This magnitude has been found appropriate since it simplifies the calibration of the experimental setup. Again higher order modes have been observed above 5 kHz for all examined cases, with the exception of the centric configurations. Below this limit planar wave propagation takes place for all configurations. These results have been confirmed in the pressure maps obtained for different formant or antiresonance frequencies. In addition, it has been observed that in the low frequency region below 5 kHz, planar waves are always generated, which only oscillate along the propagation axis and radiate following a spherical pattern. In contrast, in the high frequency region the vocal tract shape plays an important role. For centric junctions, planar waves are produced. However, when the symmetry of this geometry is broken, either with eccentric junctions or elliptical cross-sections, higher order modes appear. These modes exhibit a complex pressure distribution within the vocal tract, with a radiation pattern that can be either spherical or highly directional.

These simple examples have illustrated the importance of three-dimensional modelling of the vocal tract acoustics. On the other hand, FEM and experimental results have matched to a very good extent, so it can be concluded that the FEM approach presented in Chapter 2 can be used for vowel production simulation with confidence.

Chapter 4

Computation of vocal tract input and radiation impedances

A two-microphone transfer function method (TMTF) is adapted to a numerical framework to compute the radiation and input impedances of 3D vocal tracts of elliptical cross-section. In its simplest version, the TMTF method only requires measuring the acoustic pressure at two points in an impedance duct and the postprocessing of the corresponding transfer function. However, some considerations are to be taken into account when using the TMTF method in the numerical context, which constitute the aim of this chapter. In particular, the importance of including absorption at the impedance duct walls to avoid lengthy numerical simulations is discussed. It is also shown how the direct impedance of plane wave propagation can be computed beyond the TMTF maximum threshold frequency, by appropriate location of the virtual microphones. Virtual microphone spacing is discussed on the basis of the so called singularity factor. Examples include the computation of the radiation impedance of vowels /a/, /i/ and /u/ and the input impedance of vowel /a/, for simplified vocal tracts of circular and elliptical cross-sections.

The contents of this chapter have been published in

- Marc Arnela and Oriol Guasch (2013), "Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method," Journal of the Acoustical Society of America, 133(6), pp. 4197–4209.
- Marc Arnela and Oriol Guasch (2012), "Adaptation of the experimental two microphone transfer function method to compute the radiation impedance of ducts from numerical simulations", 19th International Congress on Sound and Vibration (ICSV19), July 8–12, Vilnius, Lithuania.

4.1 Introduction

Radiation and input impedances of vocal tracts are of special interest for voice production. They play a significant role in determining wave radiation at the lip termination or in modelling the acoustic coupling of the vocal tract and the vocal folds. Impedances can be computed from numerical simulations of vocal tract acoustics. Due to intricate geometry of the vocal tract the finite element method (FEM) is an appropriate numerical approach to carry out these simulations. Several works can be found in literature dealing with FEM computations both in the frequency domain (e.g., Matsuzaki *et al.*, 2000; Motoki, 2002; Hannukainen *et al.*, 2007) and time domain (e.g., Švancara and Horáček, 2006; Vampola *et al.*, 2008a, 2011). Occasionally other approaches such as finite differences have also been used (e.g., Takemoto *et al.*, 2010).

With regards to voice production, working in the time domain turns to be a very appealing option given that time evolving phenomena such as the self-oscillation of the vocal folds, the generation of diphthongs, or the visualization of acoustic waves propagating through the vocal tract and generation of corresponding audio files, could be naturally accounted for in numerical simulations. If the wave equation is solved in its mixed form (e.g., Takemoto *et al.*, 2010; Codina, 2008), impedances can be directly computed from the Fourier transforms of the acoustic pressure and acoustic velocity time evolutions. However, this is not possible if the wave equation is solved in irreducible form for the acoustic pressure (e.g. Vampola *et al.*, 2011) or for the velocity potential (e.g., Matsuzaki *et al.*, 2000). In such cases the acoustic velocity has to be computed from the acoustic pressure or the velocity potential gradients. This is not difficult for structured meshes but it is not so straightforward for unstructured meshes, elemental integration and assembly being required in addition at each time step for the pressure/velocity ratio at the radiation surface, in order to obtain the impedance.

In this chapter an alternative is proposed for a straightforward computation of the plane wave radiation and input impedances of vocal tracts, that requires no knowledge of the acoustic velocity field. The idea is to adapt the experimental two-microphone transfer function method (TMTF) to the numerical framework. Originally developed by Chung and Blaser (1980), the simplest version of the TMTF method only requires measuring the time evolution of the acoustic pressure at two points in an impedance duct, and computing the corresponding transfer function. From this transfer function, the radiation and/or input impedances at a given surface can be easily derived. In return, the price to be paid for such a simple approach is that an additional simulation substituting the vocal tract with an impedance duct has to be performed for the computation of radiation impedances (this is not the case for input impedances where an impedance

duct is always coupled at the entry of the vocal tract, whatever numerical approach is followed). The purpose of this study was to develop the TMTF method for computing vocal tract impedances in a numerical environment. First, it will be shown that dealing with lossy impedance ducts reduces the overall time duration of the simulations. The inclusion of wall losses allows for strong attenuation of the first duct eigenmode, which otherwise determines the total duration of the computation. However, this implies using appropriate complex wavenumbers in the TMTF expressions, which are well-known for three-dimensional circular cylindrical ducts, and which will be derived in this work for elliptical cross sectional impedance ducts, given their importance in voice production (see e.g., Motoki, 2002; Matsuzaki *et al.*, 2000, where elliptical vocal tracts are used). Second, the frequency range of validity of the TMTF method will be analyzed. Specifically, it will be shown how the direct impedance of plane wave propagation can be computed beyond the TMTF maximum threshold frequency by appropriate location of the virtual microphones (mesh nodes where the acoustic pressure time evolution is collected). In addition, the appropriate virtual microphone spacing will be determined by means of the so called singularity factor (SF) introduced by Jang and Ih (1998), according to the maximum frequency of analysis. Throughout the work, time domain FEM simulations for the irreducible wave equation will be performed with a custom software, to compute vocal tract impedances using the adapted TMTF method. However, any other time domain numerical approach could benefit from the reported results.

The chapter is organized as follows. Section 4.2 presents the methodology for computing the acoustic impedance of vocal tracts from numerical simulations. In Section 4.3, the various considerations to be taken into account when adapting the TMTF method to the numerical framework become analyzed. Numerical examples of computed impedances for vocal tracts of circular and elliptical cross-section are provided in Section 4.4. Finally, conclusions close the chapter in Section 4.5.

4.2 Methodology

4.2.1 The two-microphone transfer function method

The two-microphone transfer function method (TMTF) was originally developed by Chung and Blaser (1980) and later on standardized in the ISO 10534-2 (1998) for measuring the normal reflection coefficient of material samples. The normalized specific acoustic impedance Z can be obtained from the latter. The TMTF proceeds as follows. First, plane waves are generated at the entrance of a duct of length L, referred to as the

impedance duct. The acoustic pressure signals $P_1(f)$ and $P_2(f)$ become measured at two points x_1 and x_2 close to the impedance duct exit, which is designated as the reference surface, and the transfer function

$$H_{12}(f) = \frac{P_2(f)}{P_1(f)} \tag{4.1}$$

is computed. The reflection coefficient $\mathcal{R}_1$ at position x_1 is related to Eq. (4.1) through

$$\mathcal{R}_1 = \frac{H_{12} - H_I}{H_R - H_{12}}, \tag{4.2}$$

with H_I and H_R respectively standing for the incident and reflected wave transfer functions. Assuming plane wave propagation, the transfer functions H_I and H_R become $H_I = e^{-jk_z s}$, $H_R = e^{jk_z s}$, with $j = \sqrt{-1}$, $s = |x_1 - x_2|$ being the distance between microphones and k_z the wavenumber in the axial direction. In order to translate the reflection coefficient to the reference surface, defined at $x = 0$, a factor $e^{j2k_z x_1}$ is introduced in (4.2), leading to the following expression for the normal reflection coefficient

$$\mathcal{R} = \mathcal{R}_1 e^{j2k_z x_1} = \frac{H_{12} - e^{-jk_z s}}{e^{jk_z s} - H_{12}}\, e^{j2k_z x_1}. \tag{4.3}$$

The normalized specific acoustic impedance Z can be finally obtained by means of

$$Z = \frac{1 + \mathcal{R}}{1 - \mathcal{R}}. \tag{4.4}$$

Z is usually split in its real R (resistive) and imaginary X (reactive) components, $Z = R + jX$. On the other hand, the specific acoustic impedance Z' will be given by $Z' = Z_0 Z$, Z_0 standing for the characteristic impedance of the medium whose expression is provided in Section 4.3.2.

4.2.2 Problem statement

Suppose that the radiation impedance Z_r or the input impedance Z_{in} of a vocal tract is to be computed from a time domain numerical simulation, using the two-microphone transfer function (TMTF) method. First, an impedance duct has to be coupled to the reference surface, the duct having the same cross-sectional area and shape than this surface. For instance, to compute the input impedance of a vocal tract, the impedance duct will be coupled to the vocal tract at the glottal cross-section (see Fig. 4.1), whilst for the radiation impedance, the duct will be coupled to the mouth exit replacing the existing vocal tract geometry (see Fig. 4.1).

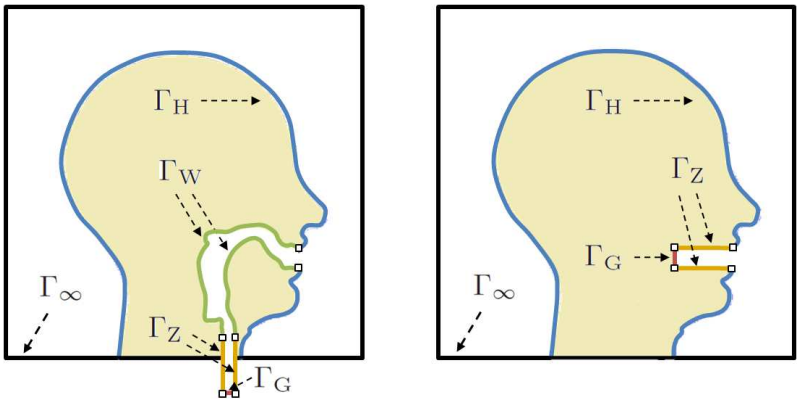


Fig. 4.1: A sketch of the computational domain Ω of Eq. (4.5) in text for input impedance computations (left) and radiation impedance computations (right). Γ_G stands for the impedance duct boundary where a volume velocity is imposed, Γ_Z for the impedance duct walls, Γ_W for the vocal tract walls, Γ_H for the human head and Γ_∞ for a fictitious non-reflecting boundary that emulates free-field conditions.

The next step consists in carrying out a time domain numerical simulation for the acoustic pressure evolution. Let us denote by Ω the finite computational domain with boundary $\partial\Omega$. $\partial\Omega$ is built from five non-intersecting regions, i.e., $\partial\Omega = \Gamma_G \cup \Gamma_Z \cup \Gamma_W \cup \Gamma_H \cup \Gamma_\infty$ with $\Gamma_G \cap \Gamma_Z \cap \Gamma_W \cap \Gamma_H \cap \Gamma_\infty = \emptyset$ (see Fig. 4.1). Γ_G stands for the impedance duct boundary where a volume velocity is imposed, Γ_Z for the impedance duct walls, Γ_W for the vocal tract walls, Γ_H for the human head boundary and Γ_∞ for the external boundary, where a fictitious non-reflecting condition has to be imposed. The problem to be solved is that of finding the acoustic pressure field $p(\boldsymbol{x}, t)$ in Ω that satisfies

$$\left(\partial_{tt}^2 - c_0^2 \nabla^2\right) p(\boldsymbol{x}, t) = 0 \qquad \text{in } \Omega,\ t > 0, \tag{4.5a}$$

with boundary conditions

$$\nabla p(\boldsymbol{x}, t) \cdot \boldsymbol{n} = -\rho_0 / S \partial_t Q(t) \qquad \text{on } \Gamma_G,\ t > 0, \tag{4.5b}$$

$$\nabla p(\boldsymbol{x}, t) \cdot \boldsymbol{n} = -\mu_w / c_0 \partial_t p(\boldsymbol{x}, t) \qquad \text{on } \Gamma_W,\ t > 0, \tag{4.5c}$$

$$\nabla p(\boldsymbol{x}, t) \cdot \boldsymbol{n} = -\mu_z / c_0 \partial_t p(\boldsymbol{x}, t) \qquad \text{on } \Gamma_Z,\ t > 0, \tag{4.5d}$$

$$\nabla p(\boldsymbol{x}, t) \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_H,\ t > 0, \tag{4.5e}$$

$$\nabla p(\boldsymbol{x}, t) \cdot \boldsymbol{n} = 1 / c_0 \partial_t p(\boldsymbol{x}, t) \qquad \text{on } \Gamma_\infty,\ t > 0, \tag{4.5f}$$

and initial conditions

$$p(\boldsymbol{x}, t) = 0,\ \partial_t p(\boldsymbol{x}, t) = 0 \qquad \text{in } \Omega,\ t = 0. \tag{4.5g}$$

In Eq. (4.5), c_0 denotes the speed of sound, ρ_0 the air density, S the impedance duct cross-sectional area at Γ_G, $\partial_t \equiv \partial/\partial t$ designates the partial time derivative and $\boldsymbol{n}$ the normal vector pointing outwards a surface. With regards to boundary conditions, $Q(t)$ in Eq. (4.5b) stands for a volume velocity generated by an imaginary loudspeaker. Eq. (4.5c) and Eq. (4.5d) account for constant frequency losses at the inner walls, being μ the boundary admittance coefficient (subindexes w and z simply indicate that different absorption values can be introduced at each boundary). μ is related to the wall impedance Z_w by means of $\mu = \rho_0 c_0 / Z_w$. Eq. (4.5e) expresses that the human head is taken as a rigid surface ($\mu = 0$). Finally, Eq. (4.5f) is the well-known Sommerfeld radiation condition, which guarantees that emanating waves from the mouth propagate outwards to infinity. However, this condition is only optimal for sound waves impinging on Γ_∞ in the normal direction. To overcome this problem, use have been made of a Perfectly Matched Layer (PML) (Berenger, 1994) for the wave equation in its irreducible form. In particular, the PML in Grote and Sim (2010) originally developed for the finite difference framework has been adapted and formulated for our custom finite element code. Details on the implemented numerical scheme can be found in Chapter 2.

As the simulation evolves, the acoustic pressure signals $p_1(t)$ and $p_2(t)$ are collected at two nodes of the finite element mesh, as if they were virtual microphones. From their Fourier Transform $P_1(f)$ and $P_2(f)$ are obtained and making use of the TMTF method described in Section 4.2.1, the acoustic impedances can be finally computed.

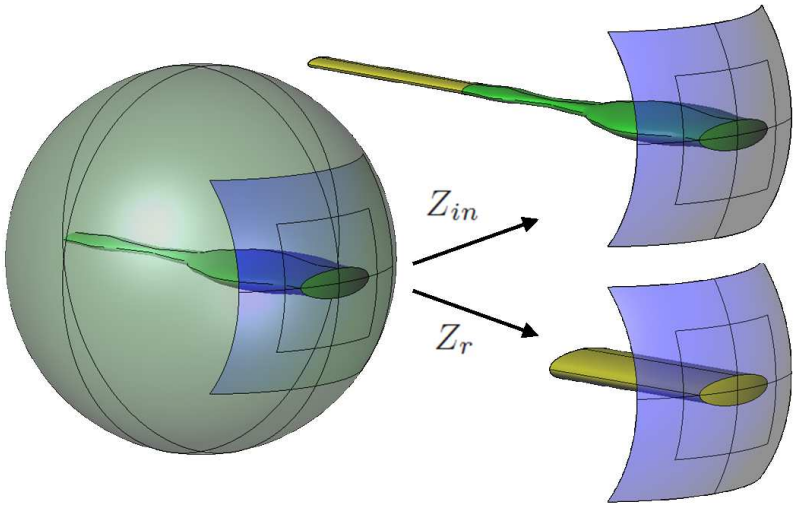


Fig. 4.2: Geometries used for computing the input impedance (right top) and the radiation impedance (right bottom) of the elliptical vowel /a/ (left).

4.2.3 Vocal tract models

The vocal tracts of the three vowels /a/, /i/ and /u/ with circular and elliptical cross-section have been considered. In order to shorten notation, these have been referred as e.g., circular /a/ or elliptical /a/. For the vocal tracts with circular cross-sections, use has been made of the simplified vocal tract geometries generated from cross-sectional areas provided by Story (2008). With regards to the vocal tracts with elliptical shapes (see Fig. 4.2), the circular cross-sections have been reshaped according to the eccentricity of the elliptical mouth apertures described in Fromkin (1964). The resulting vocal tract geometries have been set on a spherical surface of radius 0.09 m used to emulate the human head.

Following the procedure outlined at the beginning of Section 4.2.2 (see Fig. 4.2), the vocal tract geometry has been replaced by an impedance duct with equal mouth aperture to compute the vocal tract radiation impedance, whereas an impedance duct have been coupled at the glottal cross-section of the vocal tract to compute its input impedance (see Table 4.1 for duct dimensions). The impedance duct has a length of $L = 0.1$ m to fulfill the requirements of the standard ISO 10534-2 (1998) (the length should be at least three times the duct radius or the major semi-axis). The virtual microphones have been located at the centerline of the impedance duct and separated a distance $s = 0.01$ m apart (see Fig. 4.3). Following the recommendations of the ISO 10534-2 (1998), the first virtual microphone has been placed at a distance from the reference surface slightly larger than two times the impedance duct radius, or the major semi-axis. Concerning the reference surface for vocal tract radiation impedances computations, it should be noted that it is well-defined for the circular case given that the intersection of a cylindrical vocal tract with a spherical human head results in a flat disk. However, this is not the case if an elliptical vocal tract is used. In such a case, the elliptical cross-section where the major semi-axis intersects the sphere has been chosen as the reference surface (see the impedance duct exit in Fig. 4.3).

Table 4.1: Radius a, major semi-axis a_e and minor semi-axis b_e in cm of the circular and elliptical impedance ducts used to compute the radiation impedance Z_r and the input impedance Z_{in} for vowels /a/, /i/ and /u/.

	Z_r/a/	Z_r/i/	Z_r/u/	Z_{in}/a/
a	1.23	0.54	0.23	0.43
a_e/b_e	2.21/0.68	1.25/0.23	0.47/0.11	0.76/0.23

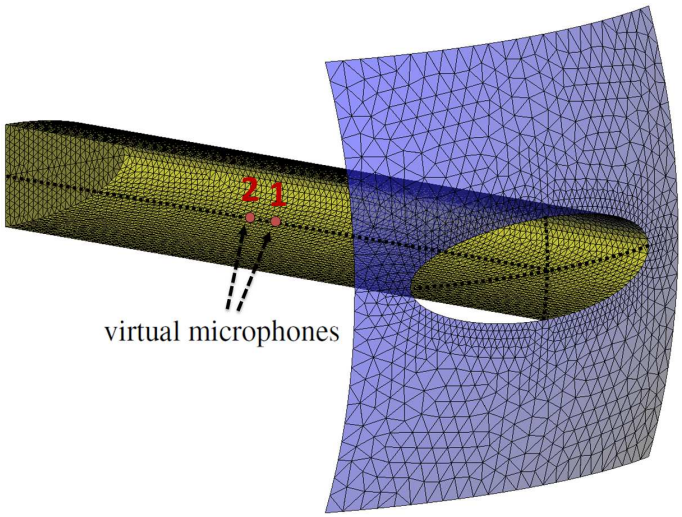


Fig. 4.3: Cut with surface mesh details of the impedance duct for the elliptical /a/ radiation impedance computation. Dots indicate locations of virtual microphones for capturing $P_1(f)$ and $P_2(f)$ in text.

4.2.4 Simulation details

The computational domain consists of an outer volume of dimensions $0.25 \times 0.2 \times 0.2$ m, where the spherical head has been placed so that sound waves can emanate from the mouth. This volume has been surrounded with a PML of width 0.1 m to absorb any incident wave. The PML has been configured to get a relative reflection coefficient of $r_\infty = 10^{-4}$ (see Chapter 2). The computational domain has been meshed using unstructured tetrahedral elements with a size comprising $h \approx 0.1$ cm inside the impedance duct, $h \approx 0.5$ cm in the outer volume and $h \approx 0.75$ cm in the PML region (see Fig. 4.3 to appreciate some mesh details).

Equation (4.5) with a PML included has been solved using the finite element approach described in Chapter 2. A time interval of total duration $T = 30$ ms has been simulated using a sampling rate of $f_s = 1/\Delta t = 2000$ kHz. The values $c_0 = 350$ m/s and $\rho_0 = 1.14$ kg/m^3 have respectively been chosen for the speed of sound and for the air density. Concerning boundary conditions, a wideband impulse has been used for the volume velocity $Q(t)$ in Eq. (4.5b), consisting of a Gaussian pulse (Takemoto *et al.*, 2010). For the boundary admittance coefficient at the vocal tract walls the value $\mu_w = 0.005$ has been used, which corresponds to the wall impedance of the vocal tract tissue $Z_w = 83666$ kg/m^{2}s (see Švancara and Horáček, 2006). For the impedance duct, the artificial value $\mu_z = 0.01$ has been chosen for the reasons explained in the following section.

4.3 The two-microphone transfer function method for numerical simulations

4.3.1 Damping the first impedance duct eigenmode

Although theoretically numerical simulations to compute vocal tract impedances could be carried out using a lossless impedance duct, including boundary losses is mandatory for the simulations to have a reasonable duration. From an experimental point of view, time duration is not a problem given that, for example, a measurement that lasts 5 seconds can be easily performed. However, in the numerical framework the CFL stability condition pose severe restrictions on the time step Δt to be used, so that for intricate and large computational domains, a 5 seconds event may involve several hours of computational time.

Consider for example, the radiation impedance computation of the circular /a/ in two cases: i) a lossless impedance duct with $\mu_z = 0$ and ii) a lossy impedance duct with $\mu_z = 0.01$. Let us first focus on the acoustic pressure collected at the first virtual microphone #1 (see Fig. 4.3) and plot its time evolution for the lossless and lossy cases in Fig. 4.4. For the former, the 30 ms duration of the simulation has not sufficed to attenuate the signal acoustic pressure inside the impedance duct, whilst it has decayed in about 15 ms for the lossy case. It should be remarked that given that the acoustic pressure has to be Fourier transformed to apply the TMTF method, it is necessary for it to vanish to zero during the simulation interval, to avoid spurious errors in this operation. Some

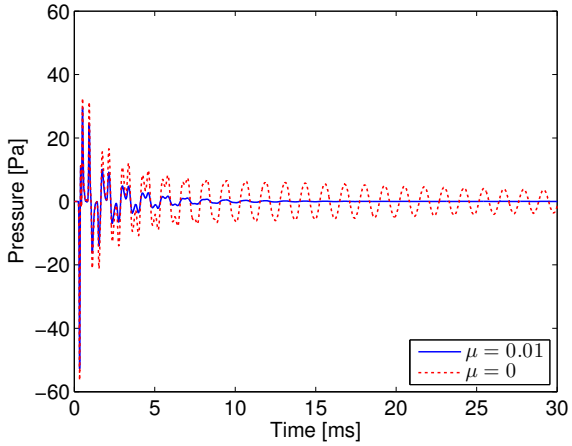
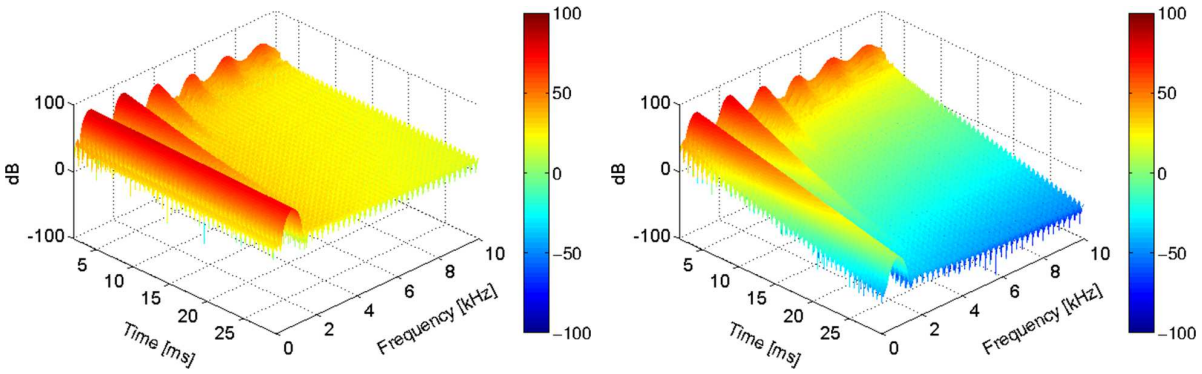


Fig. 4.4: Acoustic pressure evolution for circular /a/ at virtual microphone #1 for the lossless (μ=0) and lossy (μ=0.01) cases.



(a) $\mu = 0$ (b) $\mu = 0.01$

Fig. 4.5: Spectrograms for the acoustic pressures in Fig. 4.4: (a) lossless impedance duct (μ=0) and (b) lossy impedance duct (μ=0.01).

more insight on what is going on inside the impedance duct can be obtained from the spectrograms (time vs frequency) of the acoustic pressure signals for the lossless and lossy cases (see Fig. 4.5). As observed, the problem arises from the difficulty to attenuate the first duct eigenmode. This is attributed to the fact that radiation is by far a more effective energy dissipating mechanism at high frequencies than at low frequencies. Therefore, the inclusion of wall damping clearly helps to overcome this problem and noticeably shortens the duration of the simulation.

The value $\mu_z = 0.01$ for the impedance duct simulations of vowel /a/ has been deemed appropriate from numerical experiments. It is important to note that μ_z has to be such that the assumption of nearly hard walls is satisfied ($\mu_z << 1$). This is necessary to ensure that the mode coupling impedances are negligible (see Section 4.3.4). Besides, increasing μ_z to reduce the time duration of the computation T lowers the frequency resolution given that $\Delta f = 1/T$. Moreover, if μ_z becomes too large almost no wave is reflected from the duct termination, which results in a very poor estimate for the impedance. For instance, numerical tests with $\mu_z = 0.5$ yielded completely wrong impedances. Finally, it should be remarked that using $\mu_z = 0.01$ also for vowels /i/ and /u/ would be inappropriate. At the end of the following section it will be shown how to modify the admittance coefficient so as to guarantee that propagating waves through /i/ and /u/ impedance ducts experience the same amount of attenuation than in the /a/ duct.

4.3.2 Analytical expressions for the complex axial wavenumber

Introducing artificial damping at the impedance duct walls of the numerical model involves dealing with complex axial wavenumbers k_z in expression (4.3) of the TMTF method. In experimental measurements, a calibration procedure is usually conducted for estimating the attenuation factor (e.g. Boonen *et al.*, 2009). In the numerical case, analytical expressions can be used for k_z. k_z can be related to the admittance boundary coefficient μ_z for a generic duct of arbitrary shape with perimeter $\mathcal{P}$ and cross-section S (see e.g. Sivian, 1937). For illustrative purposes, in what follows such an expression will be derived for the particular case of three-dimensional ducts of elliptical cross-section, though following analogous reasoning lines to those in Munjal (1987) for circular and rectangular ducts.

The case of the circular duct is quite standard (see e.g., Munjal, 1987) but will be provided for completeness and for a better comprehension of the elliptical case. Assuming a harmonic time dependence $\exp(j\omega t)$ so that Eq. (4.5) transforms to the Helmholtz equation, and considering cylindrical coordinates (r, θ, z), we can express the boundary condition at the duct wall (4.5d) as

$$\frac{\partial \hat{p}}{\partial r} + j\mu_z k_0 \hat{p} = 0 \quad \text{at } r = a, \tag{4.6}$$

with $\hat{p}(\boldsymbol{x}, \omega)$ being the time Fourier transform of the sound pressure $p(\boldsymbol{x}, t)$ and $k_0 = \omega/c_0$ standing for the wavenumber. The sound pressure $\hat{p}(\boldsymbol{x}, \omega)$ at any point inside the duct admits the following series decomposition (see e.g., Munjal, 1987)

$$\begin{aligned} \hat{p}(r, z) &= \sum_{m,n=0}^{\infty} J_m\left(k_{r,mn} r\right) e^{jm\theta} \left(A_{1,mn} e^{-jk_{z,mn}} + A_{2,mn} e^{jk_{z,mn}}\right) \\ &\simeq J_0\left(k_{r,00} r\right) \left(A_{1,00} e^{-jk_{z,00}} + A_{2,00} e^{jk_{z,00}}\right), \end{aligned} \tag{4.7}$$

where J_i correspond to the Bessel functions of order i, A_i are constant factors, and $k_{z,mn}$ and $k_{r,mn}$ respectively stand for the axial and radial wavenumbers. In the second equality of Eq. (4.7) we have assumed axisymmetry ($m = 0$) and plane wave propagation ($n = 0$). Besides, note that

$$\frac{\partial \hat{p}(r, z)}{\partial r} \simeq -k_{r,00} J_1\left(k_{r,00} r\right) \left(A_{1,00} e^{-jk_{z,00}} + A_{2,00} e^{jk_{z,00}}\right), \tag{4.8}$$

so that evaluating Eq. (4.7) and Eq. (4.8) at $r = a$ and substituting in the boundary condition (4.6) yields

$$\frac{k_{r,00} J_1\left(k_{r,00} a\right)}{k_0 J_0\left(k_{r,00} a\right)} = j\mu_z. \tag{4.9}$$

For wavelengths clearly larger than the duct radius ($k_{r,00}a \ll 1$), the Bessel functions can be approximated to first order yielding $J_0(k_{r,00}a) \simeq 1$, $J_1(k_{r,00}a) \simeq k_{r,00}a/2$. Substitution in Eq. (4.9) then provides

$$k_{r,00} \simeq \sqrt{j\frac{2\mu_z k_0}{a}}. \tag{4.10}$$

Therefore, we can finally obtain the axial wavenumber $k_{z,00}$ for plane wave propagation in an axisymmetric duct through

$$k_{z,00} = \sqrt{k_0^2 - k_{r,00}^2} \simeq k_0\sqrt{1 - j\frac{2\mu_z}{k_0 a}}. \tag{4.11}$$

Next, an analogous expression to Eq. (4.11) for a duct with elliptical cross-section will be derived. To proceed, it is convenient to express the wave equation in elliptic cylindrical coordinates (ξ, η, z) (see e.g., Lowson and Baskaran, 1975). For each constant value of z, the coordinate lines correspond to confocal ellipses and hyperbolae. Curves of constant ξ are ellipses whilst curves of constant η are hyperbolae. Assuming a harmonic time dependence $\exp(j\omega t)$, so that Eq. (4.5) transforms to the Helmholtz equation, results in the following expression for the boundary condition Eq. (4.5d) at the duct wall ($\xi = \xi_0$), in elliptic cylindrical coordinates (see e.g., Oliveira and Gil, 2010)

$$\frac{\partial \hat{p}}{\partial \xi} + j\mu_z\kappa\sqrt{1 - e^2\cos^2\eta}\,\hat{p} = 0 \quad \text{at } \xi = \xi_0. \tag{4.12}$$

$\hat{p}(\boldsymbol{x}, \omega)$ stands for the time Fourier transform of the sound pressure $p(\boldsymbol{x}, t)$ and $e = \sqrt{1 - b_e^2/a_e^2}$ is the eccentricity of the ellipse defining the duct boundary at ξ_0, which has focal distance $\ell = a_e e$ and major and minor semi-axes $a_e = \ell\cosh\xi_0$ and $b_e = \ell\sinh\xi_0$. $\kappa \equiv k_0 a_e$ is the reduced (adimensional) wavenumber. To simplify notation, use will be also made of the parameter

$$q \equiv \left(\frac{k_\perp \ell}{2}\right)^2 = \left(\frac{k_\perp a_e e}{2}\right)^2, \tag{4.13}$$

with $k_\perp \equiv \sqrt{k_0^2 - k_z^2}$ standing for the transverse wavenumber.

Separation of variables for the Helmholtz equation in elliptic cylindrical coordinates results in the so-called Mathieu radial and angular equations. The solutions need to be 2π periodic in η, which plays the role somehow analogous to θ for the circular case (developed e.g. in Munjal, 1987, Chap. 1, pp. 35). The periodic solutions are given by products of cosine elliptic functions $ce_m(\eta, q)$, which are even, with the radial Mathieu functions $Je_m(\xi, q)$ related to them, and by sine elliptic functions $se_m(\eta, q)$, which are odd, and their corresponding radial Mathieu functions $Jo_m(\xi, q)$ (see Gutiérrez-Vega, 2000). However,

no linear combination of $ce_m(\eta,q)Je_m(\xi,q)$ and $se_m(\eta,q)Jo_m(\xi,q)$ is allowed since the sets of characteristic values for $ce_m(\eta,q)$ and $se_m(\eta,q)$ are different (the situation is somehow similar to what occurs for rectangular ducts, see e.g., Munjal, 1987, Chap. 1, pp. 28). Given that our interest is in plane wave propagation inside the duct, the lowest modal indices from the even case will have to be considered.

The even radial Mathieu functions $Je_m(\xi,q)$ admit a series factorization in terms of Bessel functions. The factorization depends on m being even or odd which results in even-even $Je_{2k}(\xi,q)$ and even-odd $Je_{2k+1}(\xi,q)$ radial Mathieu function developments. For the case $m=0$ the former has to be considered. It turns that (Gutiérrez-Vega, 2000)

$$Je_{2k}(\xi,q)=\frac{ce_{2k}(0,q)}{A_0}\sum_{i=0}^{\infty}A_{2j}J_{2j}(2\sqrt{q}\sinh\xi) \tag{4.14}$$

with J_m corresponding to the Bessel function of order m. Taking $k=i=0$ yields

$$\hat{p}(\xi,\eta,z)\simeq ce_0(\eta,q)ce_0(0,q)J_0\left(2\sqrt{q}\sinh\xi\right), \tag{4.15}$$

with derivative

$$\frac{\partial\hat{p}(\xi,\eta,z)}{\partial\xi}\simeq -ce_0(\eta,q)ce_0(0,q)2\sqrt{q}\cosh\xi J_1\left(2\sqrt{q}\sinh\xi\right). \tag{4.16}$$

Substituting (4.15) and (4.16) into the boundary condition (4.12) and evaluating at $\xi=\xi_0$ results in

$$\frac{2\sqrt{q}\cosh\xi_0 J_1\left(2\sqrt{q}\sinh\xi_0\right)}{\kappa J_0\left(2\sqrt{q}\sinh\xi_0\right)}=j\mu_z\sqrt{1-e^2\cos^2\eta}. \tag{4.17}$$

Integrating at both sides from 0 to $\pi/2$ allows to eliminate the η dependence in the r.h.s of Eq. (4.17). Changing variables to $v=\cos\eta$ shows that the integral in the r.h.s is a complete elliptic integral of the second kind, whose solution can be expressed as a series in terms of even powers of the eccentricity (see Abramowitz and Stegun, 1970)

$$\begin{aligned}\int_0^{\pi/2}\sqrt{1-e^2\cos^2\eta}d\eta&=\int_0^1\frac{\sqrt{1-e^2v^2}}{\sqrt{1-v^2}}dv\\&=\frac{\pi}{2}\left[1-\sum_{n=1}^{\infty}\left(\frac{(2n-1)!!}{(2n)!!}\right)^2\frac{e^{2n}}{2n-1}\right]\equiv I(e)\end{aligned} \tag{4.18}$$

Using this result, approximating the Bessel functions to first order, $J_0(x)\simeq 1$, $J_1(x)\simeq x/2$, and considering the notation introduced above, it follows from Eq. (4.17) that

$$q=\frac{j\mu_z\kappa I(e)}{\pi\cosh\xi_0\sinh\xi_0}=j\frac{\mu_z k_0 I(e)a_e^2e^2}{\pi b_e}. \tag{4.19}$$

From (4.13) it follows that

$$k_\perp^2 = j\frac{4\mu_z k_0 I(e)}{\pi b_e} \tag{4.20}$$

and the expected axial wavenumber becomes

$$k_z = \sqrt{k_0^2 - k_\perp^2} \simeq k_0\sqrt{1 - j\frac{4\mu_z I(e)}{k_0\pi b_e}}. \tag{4.21}$$

In the case of zero eccentricity $I(0) = \pi/2$ and $a_e = b_e \equiv a$ so that the axial wavenumber for the circular case in Eq. (4.11) is recovered. Given that an ellipse has perimeter $\mathcal{P} = 4a_e I(e)$ and surface $S = \pi a_e b_e$ and considering the equivalents for the circle, it follows that Eqs. (4.21) and (4.11) are nothing but particular cases of the more general expression

$$k_z \simeq k_0\sqrt{1 - j\frac{\mu_z\mathcal{P}}{k_0 S}}. \tag{4.22}$$

On the other hand, note that to compute the impedance of vocal tracts, impedance ducts of different cross-sections and shapes will be necessary. This implies that the total area of the impedance duct walls will be different in each case, so that if the same boundary admittance coefficient μ_z is used, different internal dissipation would result. To guarantee the same amount of absorption, it must be ensured that the axial wavenumber remains the same for all cases. Suppose that the same amount of dissipation is wanted for an elliptical impedance duct m than for another elliptical impedance duct n, with respective eccentricities e_m, e_n, and semi-minor axes $b_{e,m}$, $b_{e,n}$. Equating their axial wavenumbers (4.21) it is possible to compute $\mu_{z,m}$ from $\mu_{z,n}$ as

$$\mu_{z,m} = \mu_{z,n}\frac{I(e_n)}{I(e_m)}\frac{b_{e,m}}{b_{e,n}}. \tag{4.23}$$

As mentioned in Section 4.3.1, in this work a boundary admittance coefficient of $\mu_z = 0.01$ has been chosen for the circular vowel /a/. Therefore, to introduce the same amount of dissipation in all simulations, use has been made of (4.23), with $n =$/a/ and $m =$/i/,/u/.

Finally, using the complex axial wave number derived in Eq. (4.22), the characteristic impedance Z_0 in the impedance duct can be readily obtained from the ratio between the pressure and particle velocity in a lossy duct without mean flow in Munjal (1987), Chap. 6, pp. 232,

$$Z_0 = \rho_0 c_0\frac{k_0}{k_z} \simeq \rho_0 c_0\left(1 - j\frac{\mu_z\mathcal{P}}{k_0 S}\right)^{-1/2}. \tag{4.24}$$

For large ducts with small wall losses the usual value $Z_0 \approx \rho_0 c_0$ is recovered (see e.g., Dalmont *et al.*, 2001).

4.3.3 Accuracy of complex axial wavenumbers

Several assumptions have been made in order to derive the axial complex wavenumbers in Eqs. (4.21) and (4.11). Let us next check if these simple expressions are precise enough for vocal tract impedance computations and what would happen if real wavenumbers were considered instead. To do so, the radiation impedance of the circular and elliptical /i/ using complex and real wavenumbers have been computed considering different pairs of virtual microphones, located at different distances from the duct exit. The computed radiation impedances, splitted in terms of resistance R_r and reactance X_r, have been

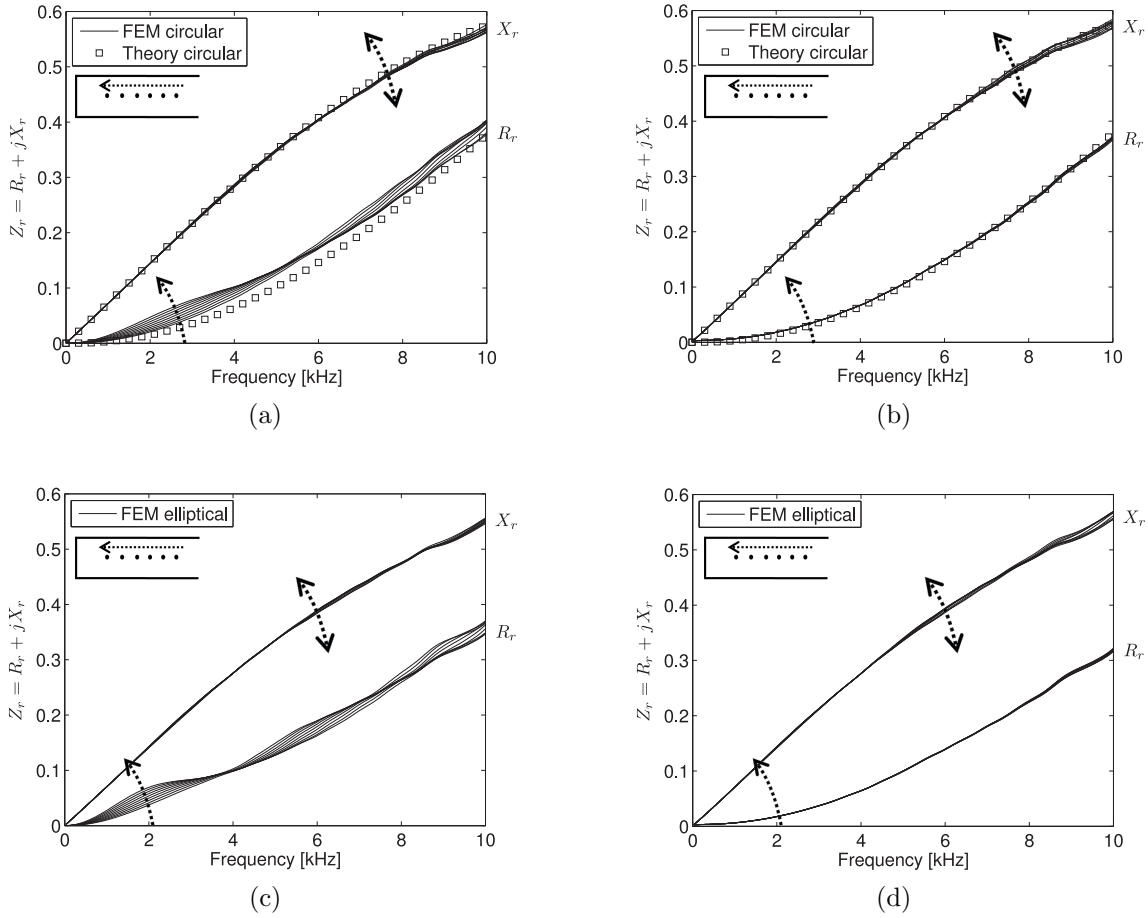


Fig. 4.6: Radiation resistance R_r and reactance X_r for vowel /i/ with circular (top) and elliptical (bottom) mouth aperture for different pairs of virtual microphones. (a) and (c) Real wavenumber computations. (b) and (d) Complex wavenumber computations. Solid lines correspond to FEM simulations and square symbols to the theoretical model, only available for a circular piston.

plotted in Fig. 4.6 and compared to the theoretical model of the baffle set in a spherical surface (Morse and Ingard, 1968, Chap. 7, pp. 343), for the circular case. The single point arrow on the R_r curves indicates their tendency when the selected pairs of microphones used for the computations are moved away from the duct exit (see duct scheme in the upper left side of figures). The double point arrow indicates that the X_r curves do not follow any clear tendency when changing the pairs of virtual microphones.

Having a look at Eq. (4.3), it can be observed that the term $e^{j2k_z x_1}$ is used to translate the reflection coefficient R_1 to the reference plane (x_1 is the distance from the first virtual microphone to the reference plane). Taking into account the first order approximation $(1+x)^{1/2} \sim 1 + x/2$ in the expression for the complex wavenumber in (4.21) (which is valid for hard wall behavior $|Z_w| \gg \rho_0 c_0$), it follows that

$$k_{z,00} \equiv k_z \simeq k_0 - j2\mu_z I(e)/(\pi b_e). \tag{4.25}$$

Substituting into the propagator $e^{j2k_z x_1}$ results in

$$e^{j2x_1 k_0} e^{j2x_1 \mu_z I(e)/(\pi b_e)}, \tag{4.26}$$

so that the factor $e^{j2x_1\mu_z I(e)/(\pi b_e)}$ will be missing if a real wavenumber is used. Note that the higher the value of x_1 the larger will be the error. This can be appreciated in Fig. 4.6a for the circular case, where R_r clearly departs from the theoretical curve as x_1 increases. As opposite, the error is inexistent when the appropriate complex wavenumber of Eq. (4.11) is used in (4.3), and the computed curve perfectly matches the theoretical one (see Fig. 4.6b). The reactance X_r seems not to be as much affected as the resistance R_r by the error.

A very similar behavior can be observed for the elliptical case as shown in Fig. 4.6c and Fig. 4.6d. Besides, note that both for the circular and elliptical cases, some discrepancies are found for the reactance values at the high frequency range, even when using the correct complex wavenumbers. This is attributed to numerical errors in the simulations to be discussed in Section 4.3.5.

4.3.4 Plane wave propagation restriction

At this point, it is convenient to comment that by using the TMTF method described in this work only the direct impedance of the planar wave mode can be computed. This is the one that dominates below the first cutoff frequency f_c of the impedance duct. However, above f_c higher order modes get excited and combine with the planar one resulting in complicate pressure patterns inside the duct (see e.g., Motoki, 2002). To

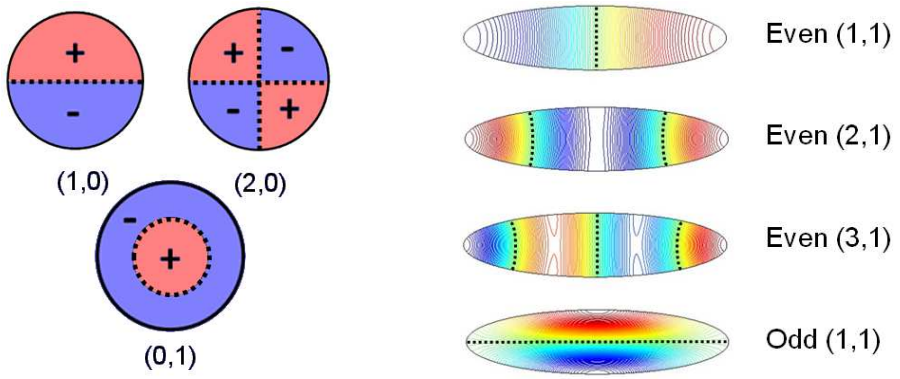


Fig. 4.7: Plots of the lower standing modes in an impedance duct with (a) circular and (b) elliptical cross-section. Discontinuous lines represent nodal planes.

capture this acoustic multi-modal behavior a direct impedance is required for each mode, leading to a matrix representation of the impedance, as described in Zorumski (1973). Coupling between modes may also occur and perturb the direct impedances. This effect is quantified by the coupling impedances and can be neglected for nearly hard walls, (see Morfey, 1969; c.f. Zorumski, 1973). This is the case for vocal tract acoustics given that the values $\mu_w = 0.005$ and $\mu_z = 0.01$ have respectively been used for the vocal tract walls and for the impedance duct walls, as detailed in the preceding sections.

Although high order modes direct impedances are needed for an accurate characterization of the radiation impedance above f_c, several theoretical models such as the piston set in a spherical baffle, which only consider the direct impedance of the plane wave propagation, are still used in 3D simulations to emulate radiation (e.g., Zhou *et al.*, 2008; Vampola *et al.*, 2008a). This is so because the computational domain (and hence the time duration of the simulation) can be strongly reduced given that no radiation space is required. Therefore, it would be interesting to use the TMTF method for computing the planar mode direct impedance above f_c, to allow comparison with those theoretical models. It will be next shown how this can be done through appropriate location of the virtual microphones.

For a circular impedance duct of radius a, the first non-planar eigenmode is the (1,0) mode, with a cutoff frequency $f_{(1,0)} = 1.84c_0/(2\pi a)$ (Fletcher and Rossing, 1988). The next one, is the (2,0) mode, with $f_{(2,0)} = 3.05c_0/(2\pi a)$, followed by the (0,1) mode, with $f_{(0,1)} = 3.80c_0/(2\pi a)$. In the experimental framework, the (1,0) mode limits the working frequency range of the TMTF method. However, in the numerical framework, this limitation can be overcome by a proper location of the virtual microphones. Having a look at the pressure distribution of the first three eigenmodes (see Fig. 4.7a), nodal planes

Table 4.2: Frequency values in kHz for the first eigenmodes of the circular and elliptical impedance ducts used to compute the radiation impedance Z_r and the input impedance Z_{in} for vowels /a/, /i/ and /u/. Stars denote maximum frequency of analysis when the virtual microphones are located at the centerline of the impedance duct.

		Circular			Elliptical			
		(1,0)	(2,0)	(0,1)	E(1,1)	E(2,1)	E(3,1)	O(1,1)
Z_r	/a/	8.33	13.81	17.2*	4.68	8.6*	12.47	13.79
	/i/	19.05	31.58	39.35*	8.27	15.27*	22.18	39.3
	/u/	45.42	75.28	93.8*	22.23	40.98*	59.49	84.2
Z_{in}	/a/	24.07	39.89	49.7*	13.57	24.97*	36.2	40.03

can be observed (corresponding to lines in the duct cross-sections of the figure), which cross the centerline of the duct for the (1,0) and (2,0) modes. Consequently, locating the virtual microphones at the centerline will avoid the pressure there being unaffected by these modes. The TMTF method will be then still applicable at these frequencies and limited by the (0,1) mode, which is the first eigenmode that does not have a nodal line at the center of the duct (see Fig. 4.7a).

With regard to elliptical ducts, the frequency range can also be extended. In Fig. 4.7b the pressure patterns of the first four eigenmodes are presented for the elliptical /a/ impedance duct (these can be computed e.g., from the formulas in Oliveira and Gil, 2010). The first non-planar mode is the even mode (1,1). This mode has a nodal plane in the center, so that locating again the virtual microphones in the centerline of the impedance duct will allow to extend the analysis up to the frequency of the even mode (2,1), which is the limiting one in this case.

The first modes of the impedance ducts used in this work are shown in Table 4.2, with stars denoting in each case the first mode without a nodal plane at its center (i.e., the limiting one). As expected, the impedance duct with a strongest restriction is that of vowel /a/ (largest mouth aperture), whereas vowel /u/ (smallest mouth aperture) presents the less stringent condition. It can also be observed that working with elliptical mouth apertures results in more restrictive frequency ranges. Note however that by locating the virtual microphones at the centerline of the impedance duct, the upper frequency limit of the experimental TMTF method has been almost increased by a factor ~ 2. Except for the elliptical /a/, the desired frequency range of analysis that goes up to $f_{max} = 10$ kHz can thus be attained.

4.3.5 The singularity of the TMTF method

It is well-known that a singularity occurs in the experimental TMTF method when half the wavelength of the acoustic pressure equals the microphone spacing s, or one of its multiples. The minimum frequency that satisfies this condition is termed the *critical frequency*,

$$f_{cr} = c_0/(2s), \tag{4.27}$$

and imposes a high frequency limit $f_u < f_{cr}$ to the method. On the other hand, a minimum low frequency limit f_d also exists, because the pressure differences measured by the two microphones will be negligible if their distance apart is very small compared to the measured wavelength. Working close to both limits is not recommended and a suitable option is to take (as proposed in Jang and Ih, 1998)

$$f_d = 0.2 f_{cr} < f < 0.8 f_{cr} \ = f_u. \tag{4.28}$$

The above situation contrasts with that of numerical simulations. Suppose that a simulation up to a given frequency f_{max}, with corresponding wavelength λ_{min}, is to be carried out. In the case of unlimited computational resources at disposition, a fine enough mesh could be used to meet the standard criterion of 10 nodes per wavelength for λ_{min} (see e.g., Ihlenburg, 1998) and numerical errors would be negligible. In such situation the only frequency upper limitation would come from the plane wave restriction discussed in Section 4.3.4, so it would be reasonable to take e.g.,

$$f_d = 0 < f < f_{max} = f_{cr} = f_u \tag{4.29}$$

as the operational working frequency range. This expression (4.29) together with Eq. (4.27) provides the following possible values for the virtual microphone spacing s

$$h \leq s < 0.5\lambda_{min}. \tag{4.30}$$

Note from (4.30), that obviously the microphone spacing s cannot be smaller than the mesh size h.

However, working with very fine meshes can be unfeasible in many simulations and the 10 nodes per wavelength criterion is often sacrificed to lower the computational cost. This results in the appearance of some numerical errors at high frequencies as happens with the simulations performed throughout this work, where 34.5, 6.9 and 4.6 nodes per λ_{min} have been respectively taken within the impedance duct, the outer volume and the PML region (see Section 4.2.4 for mesh details). In Fig. 4.8 plots of the radiation resistance and reactance for the elliptical /i/ are given, computed using different microphone spacings, $s = 0.1$, 0.25, 0.5, 0.75, 1, 1.25, 1.5 cm, which respectively yield $s/\lambda_{min} \sim 0.03$, 0.07, 0.15,

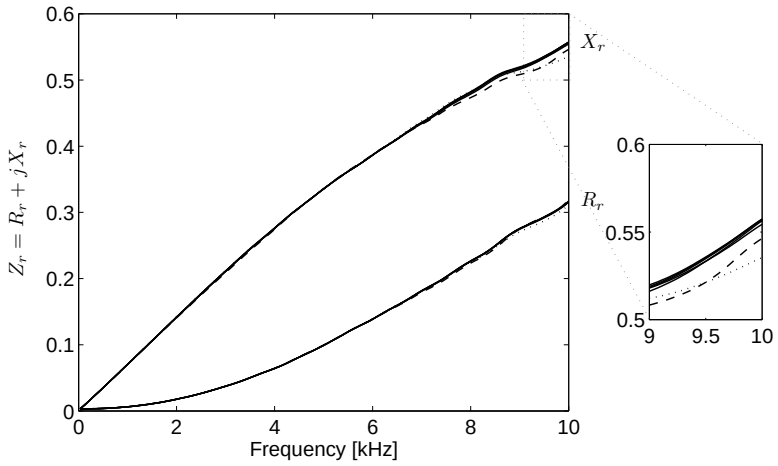


Fig. 4.8: Resistance and reactance of the elliptical /i/ for different microphone spacings s. The dashed line stands for $s = 0.1$ cm (equal to the mesh size h), the dotted line for $s = 1.5$ cm (close to the singularity), and solid lines correspond to intermediate values of s.

0.22, 0.29, 0.36, 0.43. The first spacing ($s = 0.1$ cm) has been chosen equal to the mesh size h inside the impedance duct (dashed curve in Fig. 4.8), whilst the last one ($s = 1.5$ cm) has a value close to the singularity value $0.5\lambda_{min}$ (dotted curve in Fig. 4.8). No differences can be appreciated in Fig. 4.8 between the various R_r and X_r curves for the low-mid frequency range as there are no significant numerical errors at these frequencies for the present simulations. Slight differences only become apparent for frequencies higher than 8 kHz (see zoom in Fig. 4.8), the largest ones precisely corresponding to the reactances computed with the limiting values of $s = 0.1$ cm (mesh size) and $s = 1.5$ cm (singularity value).

The above example shows that the condition in Eq. (4.30) for the microphone spacing is too loose and that stronger requirements are needed in practice for the spacing limiting values. In order to define them, resort is made to the so called singularity factor (SF) of the TMTF method, introduced by Jang and Ih (1998). The SF indicates the sensitivity of the TMTF method to errors in the input pressures $P_1(f)$ and $P_2(f)$; the higher the SF value the stronger the influence of the error sources in the computed impedances. When computing the SF in the experimental framework, it is assumed that all errors are of the white type, uncorrelated and with constant variance. This will not be the case of a single numerical simulation, the error being totally deterministic i.e., always the same if the computation becomes reran. However, considering the simulation of a given vocal tract impedance being representative of a certain ensemble average of vocal tracts having e.g., slight different geometry details and material characteristics, it is reasonable to

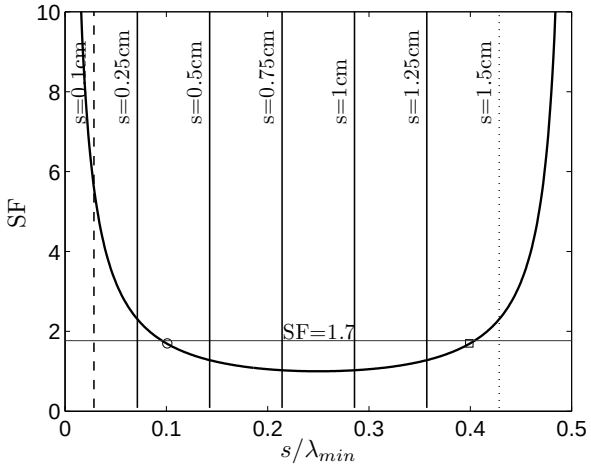


Fig. 4.9: Singularity factor SF of the TMTF method for a lossless impedance duct. The circle and the square denote the low and high limits for the microphone spacing s obtained when SF = 1.7. Vertical lines correspond to the microphone spacing configurations used in the example of Fig. 4.8.

hypothesize that errors arising from these simulations would satisfy the error requirements for the SF computation. It would then be logical to demand that the chosen spacing for the virtual microphones results in a small SF value.

In Fig. 4.9, the SF curve (thick line) for the standard TMTF method in a lossless impedance duct (implementation of the SF for a lossy impedance duct is out of the scope of this work) has been plotted according to the procedure described in Jang and Ih (1998). However, instead of representing the SF dependence with frequency, this has been plotted versus the microphone spacing s (normalized by λ_{min}). Moreover, the microphone spacings corresponding to the resistance and reactance curves in Fig. 4.8 have been also plotted as vertical lines. As observed in the figure, the SF values for the extreme spacings corresponding to the mesh size and singularity values are much higher than the threshold value of SF ≤ 1.7, recommended in the experimental framework (see Jang and Ih, 1998). According to this criterion a more restrictive range is obtained for the virtual microphone spacing than that provided by Eq. (4.30), namely

$$h' < 0.1\lambda_{min} < s < 0.4\lambda_{min}. \tag{4.31}$$

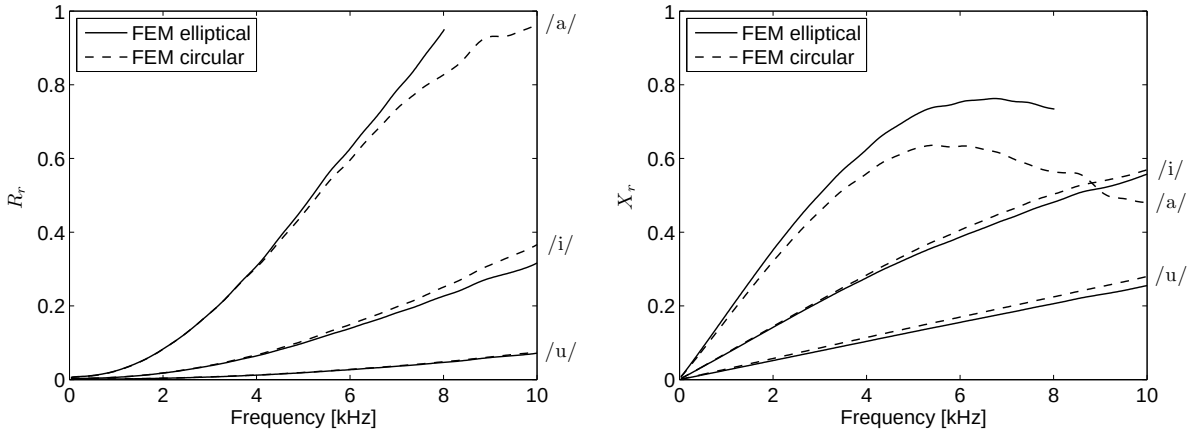
The optimum, and thus recommended, spacing s will be that minimizing SF, which is close to $s \simeq 0.25\lambda_{min}$ in Fig. 4.9. In the computations $s = 1$ cm has been used (see Section 4.2.4) that yields $s \simeq 0.29\lambda_{min}$.

4.4 Vocal tract impedances

In this section, radiation and input impedances for vowel vocal tracts using the described FEM-TMTF approach will be presented. As detailed in Section 4.2, the radiation impedance for vowels /a/, /i/ and /u/ has been computed considering circular and elliptical mouth apertures, as well as the input impedance of vowel /a/ for circular and elliptical vocal tracts.

In Fig. 4.10 the results for the radiation impedance are given, split in terms of the radiation resistance (Fig. 4.10a) and reactance (Fig. 4.10b). All results are provided up to 10 kHz except for the elliptical /a/, the analysis only being valid in this case up to $\sim$ 8 kHz because of the limiting even mode (2,1) (f = 8.6 kHz, see Table 4.2). Note however, that this value can be reached thanks to proper location of the virtual microphones (see Section 4.3.4) and that the experimental TMTF would have only let to measure impedance values up to $f = 4.68$ kHz, corresponding to the even mode (1,1) (see Table 4.2). Similarly, the plane wave frequency restrictions for the circular /a/ (f = 8.33 kHz) and for the elliptical /i/ ($f = 8.27$ kHz) in the experimental TMTF, can be easily overcome thanks to centerline microphone positioning. Thus, for all analyzed cases, except for the elliptical /a/, the radiation impedance can be computed for the whole frequency range of interest (0 10] kHz, without problems (see Table 4.2).

The differences between the resistance and reactance curves for the circular and elliptical mouth apertures in Fig. 4.10 can be justified as follows (remember that for a given vowel



(a) (b)

Fig. 4.10: (a) Radiation resistance and (b) reactance for the elliptical and circular vocal tracts of vowels /a/, /i/ and /u/.

only the shape of the mouth, elliptical or circular, changes but not the total mouth surface). In the plane wave propagation regime, the curves should be almost identical. However, as explained in Section 4.2.4, the reference surface for elliptical mouth apertures is defined at the intersection between the major semi-axis of the impedance duct with the sphere. This implies that there is a small partial protrusion of the duct beyond the reference surface, which reaches the intersection between the minor semi-axis and the sphere, and plays somehow the role of some kind of lips. Its effects on the radiation impedance of the elliptical ducts can be understood as those of a modified end correction δ. It is clear that changes in δ affect both, the resistance R and the reactance X, but for $R_r^2 + X_r^2 << 1$ only the latter becomes substantially modified. This can be viewed as follows. The reflection coefficient $\mathcal{R}$ is related to the normalized impedance $Z = R + jX$ by

$$\mathcal{R} = \frac{Z-1}{Z+1} = \frac{(R-1)+jX}{(R+1)+jX} = \frac{R^2+X^2-1}{(R+1)^2+X^2} + j\frac{2X}{(R+1)^2+X^2}, \tag{4.32}$$

its phase θ being given by

$$\tan\theta = \frac{\mathrm{Im}\mathcal{R}}{\mathrm{Re}\mathcal{R}} = \frac{2X}{R^2+X^2-1}. \tag{4.33}$$

For $R^2 + X^2 << 1$ we get $\tan\theta \simeq -2X$ and if X has small values, $\theta \simeq -2X$. The inertial correction is then given by

$$\delta = \frac{\theta - \pi}{-2k_0} = \frac{-2X - \pi}{-2k_0}, \tag{4.34}$$

showing that if $R^2 + X^2 << 1$ the end correction mainly affects the reactance. This is what can be observed for vowel /u/, which shows the same resistance values for the circular and elliptical cases (see Fig. 4.10a), but different ones for the reactance (see Fig. 4.10b). For vowels /a/ and /i/ the condition $R_r^2 + X_r^2 << 1$ is no longer satisfied, say for frequencies bigger than 4 kHz, and differences in resistance values can become more clearly appreciated.

Finally, moduli results for the input impedance of the circular and elliptical /a/ are plotted in Fig. 4.11 (results for vowels /i/ and /u/ have not been included as very similar conclusions to those of /a/ could be deemed). In contrast to the radiation impedance computation, the input impedance for elliptical /a/ can be computed with no problem up to 10 kHz. This is so because the glottal cross-section is much smaller than the mouth aperture; a narrower impedance duct is then required, which will have the first non-planar mode beyond 10 kHz. Looking at Fig. 4.11, it can be observed that below 5 kHz the input impedance moduli of the circular and elliptical /a/ are very similar indicating that they

are both in the plane wave propagation range. Note however, that there is a certain formant shift to lower frequencies for the elliptical /a/, because its radiation reactance is higher than that of the circular /a/ (see Fig. 4.10b). Above 5 kHz, differences become notorious due to non-planar high order mode effects.

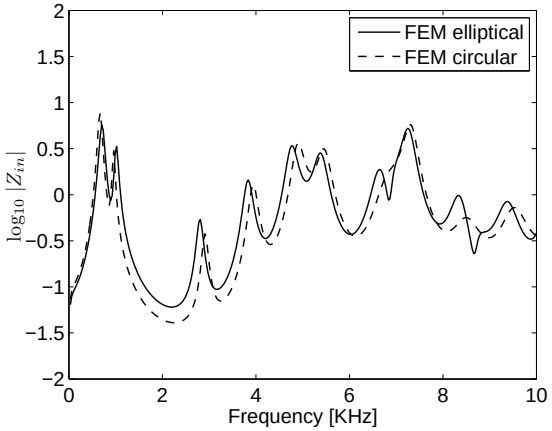


Fig. 4.11: Input impedance for the circular and elliptical /a/.

4.5 Conclusions

The radiation and input impedances of vowel vocal tracts, with circular and elliptical cross-sections, have been computed using an adaption of a two-microphone transfer function (TMTF) experimental method to a numerical framework. The procedure is rather straightforward and does not require any computation of the acoustic velocity field though, in return, it demands an additional simulation with an impedance duct to obtain radiation impedances. Despite of the simplicity of the method, some considerations are to be taken into account. First, it has been shown that it is mandatory to impose losses at the impedance duct walls to attenuate the first duct eigenmode and achieve reasonable computational times. This implies using complex wavenumbers in the TMTF expressions, which have been derived for three-dimensional cylindrical impedance ducts with elliptical cross-section. Second, the frequency range of validity of the experimental TMTF method can be almost doubled by locating the virtual microphones at the impedance duct centerline. This allows computing the direct impedance for plane wave propagation above the experimental limitation. Third, a range of possible values for the virtual microphone spacing has been proposed making use of the so called singularity factor. An optimum value of a quarter wavelength of the maximum frequency to be solved has been recommended.

Chapter 5

Analysis of the radiation effects on vowels

Radiation losses due to outward propagating waves outside the mouth are one of the most important loss mechanisms of voice generation. In this chapter, their effects on vowel production are illustrated by means of FEM simulations with free-field propagation. Two different situations are considered. In the first one, the computational domain ends at the mouth aperture, and an open end boundary condition is imposed on it, whereas in the second one free-field radiation is allowed, using a PML approach to absorb the outgoing sound waves. Differences in formant location and bandwidth are reported and related to the effects of the real (resistance) and imaginary (reactance) components of the computed radiation impedances, presenting a clear non-linear behavior due to formant cavity affiliation. Consequently, small changes in the radiation modelling can lead to large variations in vocal tract transfer functions, which highlights the importance of appropriate modelling this loss mechanism. Finally, it is shown that for limited computational domains ending at the mouth aperture, simulation results can be greatly improved by lengthening the vocal tract through appropriate end correction. This is done at almost no additional computational cost and could be useful for vowel identification as well as for medical applications.

This chapter corresponds to an elaborated version of

- Marc Arnela, Oriol Guasch and Francesc Alías (2012), "Analysis of the radiation effects on vowels by means of time domain finite element simulations," 19th International Congress on Sound and Vibration (ICSV19), July 8–12, Vilnius, Lithuania.

5.1 Introduction

Radiation loss is known to be the major loss effect of voice production in the mid and high frequency range. Three strategies have been followed to date to model it (see Fig. 5.1). In the first one (Fig. 5.1a), the computational domain is ended at the vocal tract exit (mouth) and an open-end boundary condition ($p = 0$) is simply imposed on it (see e.g., Hannukainen *et al.*, 2007). In the second one (Fig. 5.1b), the above condition is replaced by a frequency dependent impedance load (see e.g., Zhou *et al.*, 2008; Vampola *et al.*, 2008a,b). Several theoretical models for radiation impedance can be found in literature (see Appendix A), which range from Flanagan's approximation to the piston set in an infinite baffle, to the exact solution for the piston set in the infinite baffle, or to the piston set in a sphere, which better fits the head geometry (Chalker and Mackerras, 1985). In the third strategy (Fig. 5.1c), a free-field solution is adopted. This consists in extending the computational domain out of the vocal tract so sound waves can propagate from the mouth aperture towards infinity. The computational domain is truncated using for instance non-reflecting boundary conditions (see e.g., Kako and Touda, 2006), Perfectly Matched Layers (see e.g. Takemoto *et al.*, 2010; Arnela and Guasch, 2013) or infinite elements (see e.g., Švancara and Horáček, 2006; Vampola *et al.*, 2011) to avoid spurious reflections from the boundaries. This option directly accounts for radiation losses and allows us to consider any human head geometry detail, but at a higher computational cost. The alternative are the radiation load models. However, they are not straightforward to implement in the time domain due to the frequency dependence of the theoretical models. Therefore, sometimes the open-end strategy is adopted to circumvent it, but at the price of not considering any radiation loss effect.

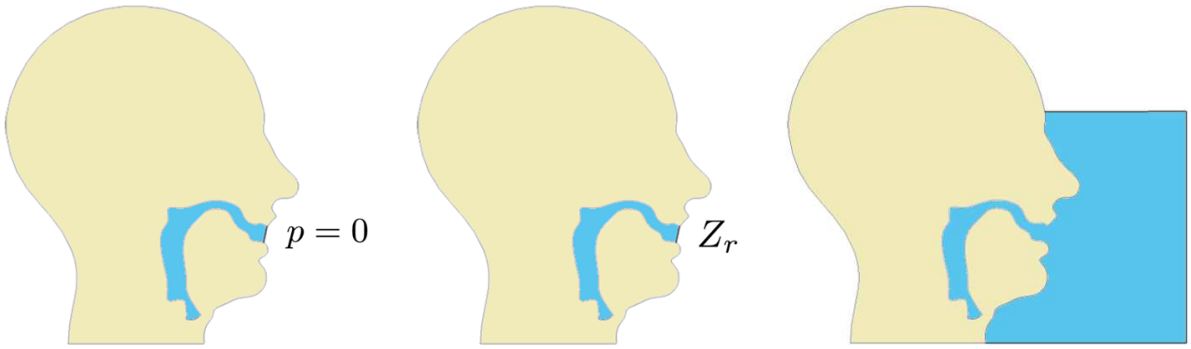


Fig. 5.1: Strategies to emulate acoustic radiation: (a) open-end boundary condition ($p = 0$), (b) radiation load models and (c) free-field radiation solutions.

In the present work, the radiation effects on vowel generation will be illustrated by contrasting a free-field radiation solution against an open-end strategy ($p = 0$). In order to facilitate comparisons with radiation load models, a spherical baffle will be used to emulate the human head, vocal tract geometries with circular cross-sections will be constructed, and the acoustic pressure will be sampled inside the vocal tract. Vowels /a/ and /u/ will be selected, since they respectively correspond to limiting cases of large and small mouth aperture, and also vowel /i/ because it has an intermediate value. A Perfectly Matched Layer (PML) approach will be adopted to emulate radiation losses, which combined with the time domain wave equation for the acoustic pressure constitutes the proposed solution for vowel production. This problem will be solved using FEM (see Chapter 2). In addition, a 1D frequency domain model that uses a spherical baffle impedance load will be included for comparisons. Note that this radiation model corresponds to the best approximation for the radiation losses that a 1D model can perform, since on the one hand no more realistic analytical models for the human head radiation impedance can be found in the literature, and on the other hand free-field radiation solutions cannot be obviously considered because they require to compute the acoustic pressure outside the vocal tract.

This chapter is structured as follows. The acoustic radiation in vowel generation will be analyzed in Section 5.2. First, the radiation impedance will be computed following the methodology described in Chapter 4 and contrasted to widely used theoretical models (Section 5.2.2). Then, the vocal tract transfer functions (Section 5.2.3) will be obtained for the free-field radiation case (PML) and compared to the open-end case ($p = 0$). Moreover, the differences with respect to the 1D approach that uses a spherical baffle impedance load will be examined. Radiation impedances and vocal tract transfer functions will then be related in Section 5.2.4 and the influence of resistance and reactance on formant location and bandwidth will be analyzed, showing a rather intricate behavior partly due to formant-cavity affiliation (Stevens, 2000). A non-standard representation for this phenomena will be given from which front and back cavity formants can be easily identified. In addition, for situations in which computational cost constitutes a limitation and/or a lower degree of precision may be admissible (e.g., medical applications or vowel identification techniques), in Section 5.3 we will show that results for simulations with computational domains ending at the mouth aperture can be greatly improved by appropriate lengthening of the vocal tract. End corrections from different impedance load models will be used and compared for this purpose. Conclusions will finally close this chapter in Section 5.4.

5.2 Influence of radiation on vowel production

5.2.1 Models and simulation details

As commented in the Introduction, vowels /a/, /i/ and /u/ are chosen to examine radiation effects on vowels. Simplified vocal tract geometries with circular cross-sections are generated using the cross-sections in Story (2008), see Fig. 5.2a. Two different cases are considered. In the first one, outward propagating sound waves from the mouth exit are allowed. In the second one, the computational domain is limited to the interior of the vocal tract and a homogeneous Dirichlet boundary condition $p = 0$ is imposed at the mouth aperture, to emulate an open-end condition. For brevity, the above approaches will be hereafter designated as the free-field radiation case and the open-end case.

Time domain finite element simulations are then performed. We will follow the computational features in Section 2.4, but with the modifications described below. In the free-field radiation case, the vocal tracts are inserted in a spherical surface of radius 0.09 m, intended as an approximation for the human head. The spherical surface is surrounded by a volume that allows emanating waves from the mouth to propagate towards infinity. To deal with a finite computational domain, a PML of width 0.2 m is used to absorb outward propagating waves (see Fig. 5.2b). An unstructured mesh of tetrahedral elements is used for all simulations. The mesh size h ranges from $h \approx 0.001$ m inside the vocal tract to $h \approx 0.01$ m at the outer PML region. Numerical simulations are performed with a sampling rate of $f_s = 1600$ kHz and a duration of $T = 50$ ms. The values $c_0 = 345$ m/s and $\rho_0 = 1.1933$ kg/m^3 are chosen respectively for the speed of sound and air density.

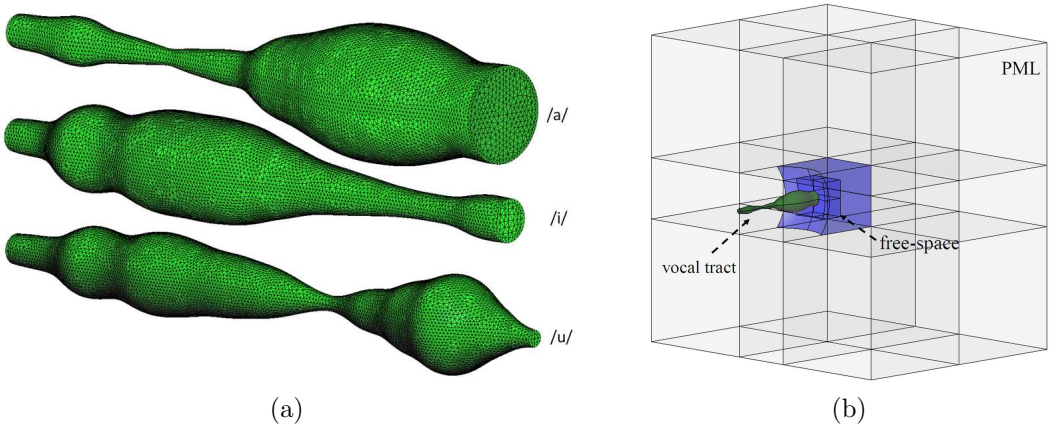


Fig. 5.2: (a) Surface mesh details for vowels /a/, /i/ and /u/ used in the open-end case. (b) Computational domain for vowel /a/ used in the free-field radiation case.

5.2.2 Computing the radiation impedance of vowels

The specific acoustic radiation impedances Z_r of vowels /a/, /i/ and /u/ is first computed for the free-field radiation case. An adaptation of the two-microphone transfer function method to the numerical framework (see Chapter 4) is used to numerically compute the real R_r (resistance) and imaginary X_r (reactance) parts of Z_r, for every vowel. Results are shown in Fig. 5.3. We also include in this figure the theoretical model of a piston set in a sphere to validate the FEM simulations, the piston set in an infinite baffle and the well-known Flanagan's approximation (see Appendix A). Comparing the FEM results with the spherical baffle ones, we observe that the simulated resistance and reactance for each

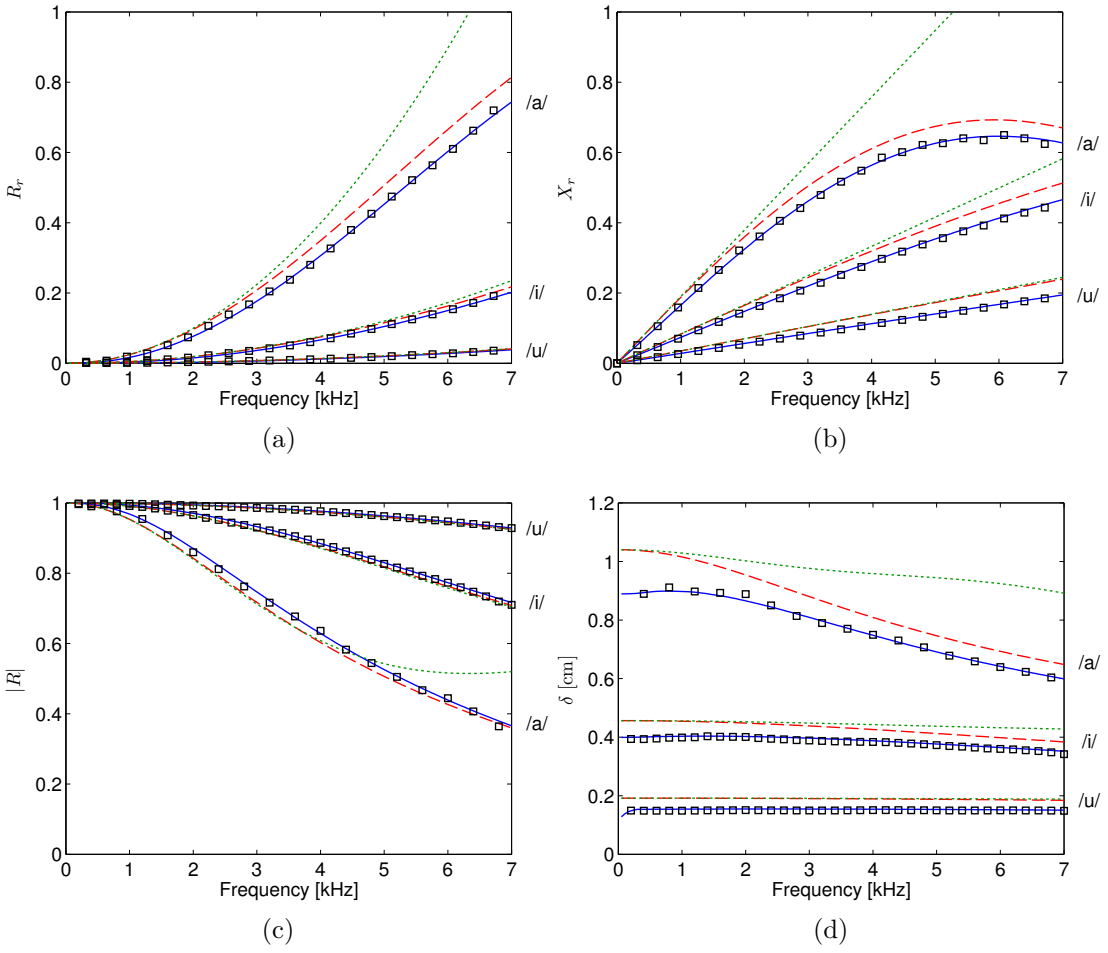


Fig. 5.3: Acoustic radiation of vowels /a/, /i/ and /u/: (a) Normalized resistance R_r, (b) normalized reactance X_r, (c) reflection coefficient modulus $|R|$ and (d) inertial end correction δ. (□ FEM, — Spherical baffle, - - Infinite baffle, · · · Flanagan).

vowel fit with accuracy the theoretical curves for the whole frequency range, regardless of the mouth aperture. The same holds true for the related variables, reflection coefficient and end correction length.

On the other hand, large differences can be appreciated when contrasting the infinite baffle and Flanagan results in Fig. 5.3, with the spherical model (or the FEM one), specially for large mouth apertures and high frequency values. Moreover, given that Flanagan's derivation is an approximation to the infinite baffle for $ka << 1$ (a being the mouth radius) differences can be also appreciated between them, specially for vowels /a/ and /i/, for frequencies above 2 kHz and 1 kHz respectively. However, no significant deviations are found for vowel /u/. The implications of such behavior will be discussed to some extent in Section 5.3. Note that the results in Fig. 5.3 state the obvious fact that the radiation impedance increases with frequency and mouth aperture. With regard to the generation of vowels, one could infer that radiation effects will be mainly of importance for such regimes. It will be shown in Section 5.2.4 that this is not always the case.

5.2.3 Computing the vocal tract transfer functions

Next, time domain finite element simulations as described in previous Section 5.2.1 are performed for the open-end and free-field radiation cases. A snapshot at time instant $t \approx 1$ ms of the acoustic pressure for vowel /a/ is presented in Fig. 5.4. The vocal tract transfer function (VTTF) between the acoustic pressure P_o at the mouth and the volume

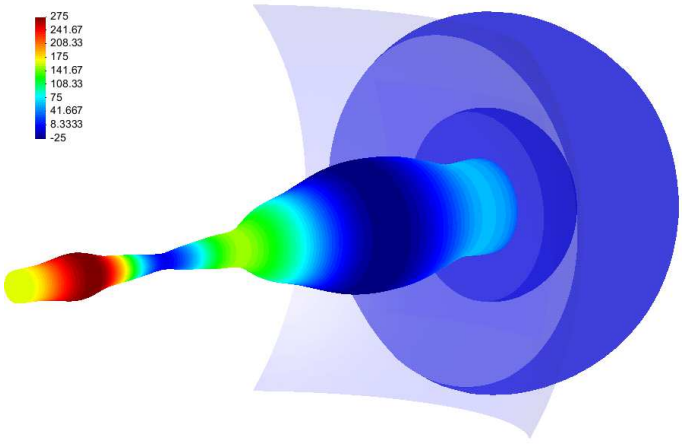


Fig. 5.4: Acoustic pressure field (Pa) of vowel /a/ at time instant $t \approx 1$ ms for the free-field radiation case.

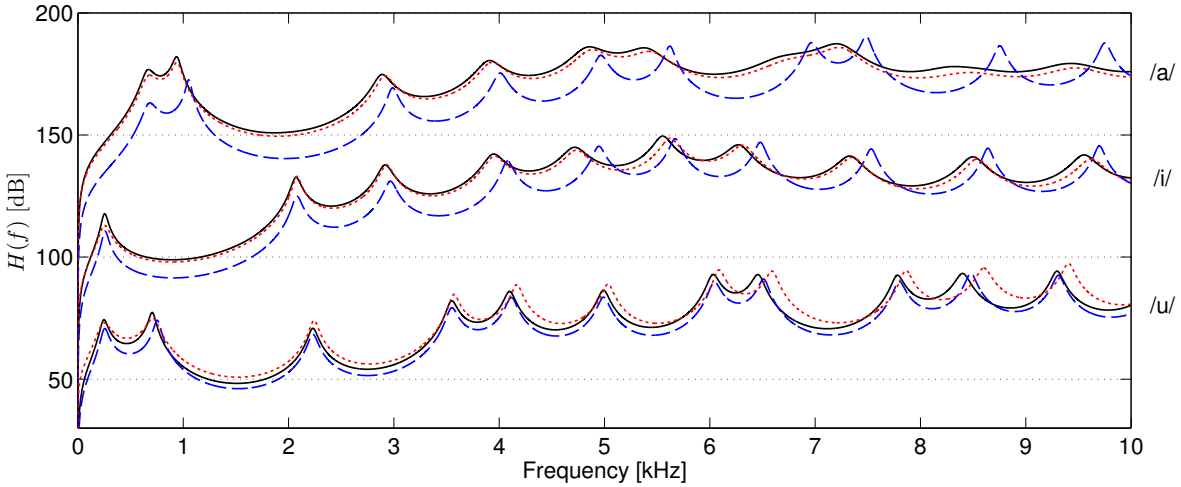


Fig. 5.5: Transfer function $H(f) + 50$ dB for vowel /a/, $H(f)$ for vowel /i/ and $H(f) - 50$ dB for vowel /u/. (— Free-field radiation (3D-FEM), - - Open-end (3D-FEM), · · · 1D-Spherical).

velocity Q_i (a Gaussian pulse) input at the glottal folds boundary,

$$H(f) = \frac{P_o(f)}{Q_i(f)}, \tag{5.1}$$

are computed. P_o is taken from a node inside the vocal tract located at 0.3 cm from the exit to allow fair comparison between the free-field radiation and open-end cases. The corresponding VTTFs are shown in Fig. 5.5. For comparison purposes, we also include the VTTFs of a 1D model using the impedance load of a piston set in a spherical baffle (referred as 1D-Spherical). They have been computed using standard chain matrices for acoustic transmission lines (see e.g., Fahy, 2001), with the same wall losses than for the FEM case. As expected, large differences between the open-end and free-field radiation cases can be observed. When free-field radiation is allowed, the formant locations are shifted down, their bandwidths (3dB decay in Hz) increase, and their amplitudes become modified. In Table 5.1, the first four ($i = 1, 2, 3, 4$) values for formant locations F_i and bandwidths BW_i for the open-end and free-field radiation cases are summarized. Moreover, the differences between them (Δ) are also reported. If we next focus on the 1D-Spherical case (see Fig. 5.5), in contrast to the free-radiation case (3D-FEM), in this model non-planar waves can not be considered, so that significant differences are produced specially for frequencies above 4 kHz. This results for instance that for vowel /u/ better results are obtained for a 3D-FEM approach with an open-end condition than for a 1D solution with a spherical baffle.

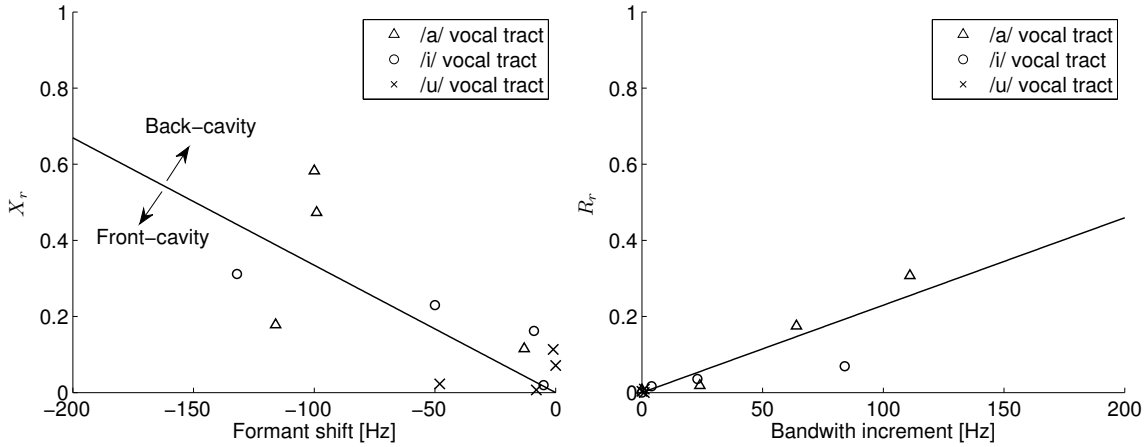
Table 5.1: Formant location F_i (Hz) and bandwidth BW_i (Hz) computed for the first four formants ($i = 1, 2, 3, 4$) of vowels /a/, /i/ and /u/. The results are obtained for the open-end (Open) and free-field radiation (Rad) cases. Their differences are computed in Δ.

	F_1			F_2			F_3			F_4		
	Open	Rad	Δ	Open	Rad	Δ	Open	Rad	Δ	Open	Rad	Δ
/a/	684	671	-13	1053	937	-116	2989	2890	-99	4011	3911	-100
/i/	259	254	-5	2081	2072	-9	2968	2918	-50	4081	3949	-132
/u/	256	248	-8	754	706	-48	2232	2232	0	3552	3551	-1

	BW_1			BW_2			BW_3			BW_4		
	Open	Rad	Δ	Open	Rad	Δ	Open	Rad	Δ	Open	Rad	Δ
/a/	133	133	0	50	74	24	85	149	64	116	227	111
/i/	57	57	0	80	84	4	99	122	23	96	180	84
/u/	69	68	-1	61	62	1	84	84	0	93	94	1

5.2.4 The radiation impedance versus the transfer functions

Finally, we discuss the radiation effects in vowels in terms of radiation impedance (see Section 5.2.2) and vowel formants (see Section 5.2.3). It is well-known that the main roles played by the reactance X_r and the resistance R_r are respectively those of frequency shifting and damping of the vowel formants. In Fig. 5.6, a plot is made of X_r and R_r (see Fig. 5.3) versus the formant frequency shift and bandwidth increment (see Table 5.1). Moreover, as a reference, the figure also includes the same results for constant cross-section tubes (a linear relationship is obtained in this case). We can observe that opposite to constant cross-section tubes, there is no linear relationship between the reactance X_r and the frequency shift (see Fig. 5.6a). For instance, if we have a look at the results for vowel /u/ (see Table 5.1), it is apparent that there is no significant deviation for any formant except for the second one, which reaches the substantial value of -48 Hz. In what concerns the resistance R_r, it is clear from Fig. 5.6b that its dependence with bandwidth is also no longer linear. On the other hand, note from Table 5.1 that the first formant of all vowels is not affected by radiation damping. This confirms that radiation losses are of importance mainly in the mid-high frequency range. The above non linear behaviors between radiation impedance and radiation effects can be understood by means of the formant-cavity affiliation concept (see Fant, 1970), where each formant is interpreted as a resonance of a specific vocal-tract cavity. For example, for vowel /u/, the second formant corresponds to a front-cavity whilst the first, third and fourth formants correspond to



(a) (b)

Fig. 5.6: Effects of resistance and reactance on vocal tract transfer functions: (a) Formant shift versus reactance and (b) bandwidth increasing versus resistance. The solid line stands for the behavior of constant cross-section ducts, and delimits front-cavity formants from back-cavity formants.

back-cavities (see Apostol *et al.*, 2004). To illustrate it, in Fig. 5.7 the acoustic pressure distribution within the vocal tract is shown for the second and third formants of vowel /u/, for the open-end case. Simulations have been carried out according to Section 5.2.1, but using a pure tone with equal frequency to the formant of interest instead of the Gaussian pulse. For the second formant, we can observe that a much higher pressure values are achieved at the front-cavity than in the back-cavity. In contrast, for the third formant the highest pressure levels are located in the back-cavity. Front-cavity resonances are strongly influenced by acoustic radiation, whilst the latter's effect on back-cavity resonances is much weaker (Stevens, 2000). This explains why radiation effects are noticeable for the second formant of vowel /u/, yielding to the strong frequency shift value of -48 Hz (see Table 5.1), whereas no changes can be appreciated for the remaining formants. Given that constant cross-section tubes consist of a single cavity, all resonances become proportionally influenced by radiation, which results in the linear behavior observed in Fig. 5.6. Moreover, this straight line for the constant cross-section duct's behavior establishes a boundary between front-cavity (below the line) and back-cavity (above the line) formants. Therefore, we have observed that by means of the analysis of the reactance effects the formant cavity affiliation could be determined.

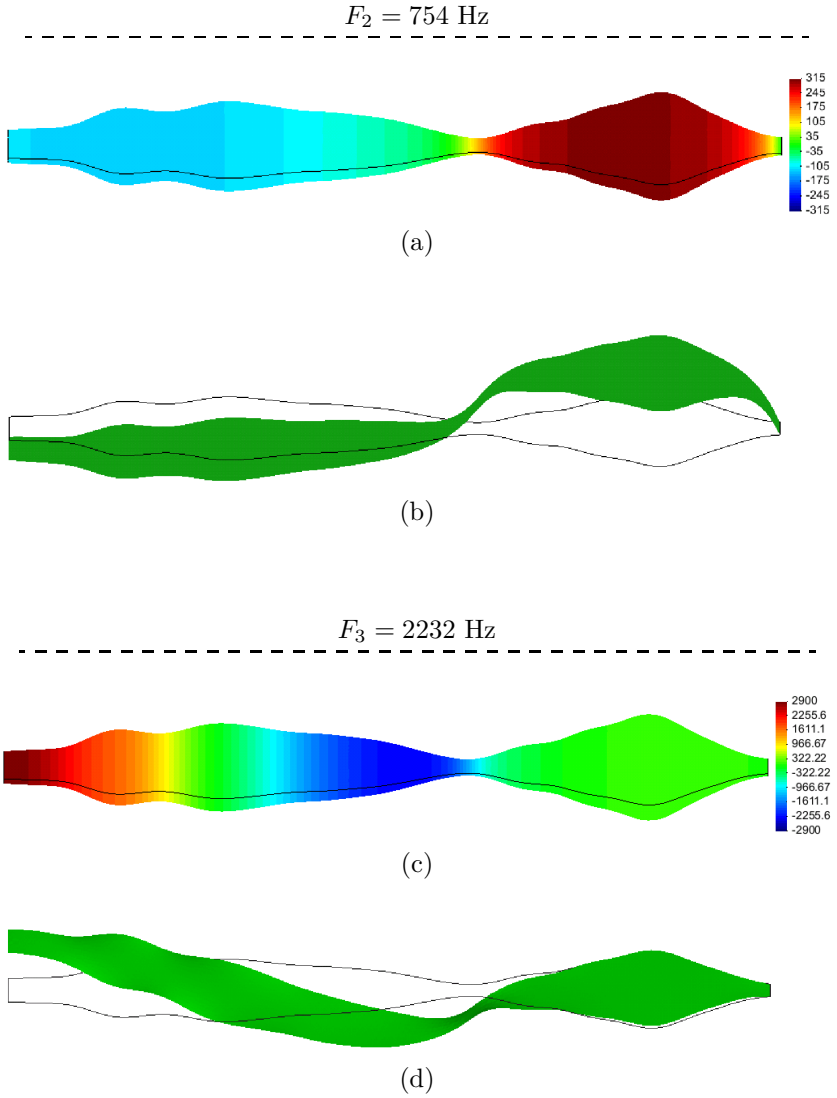


Fig. 5.7: Acoustic pressure distribution for the (a-b) second and (c-d) third formant of vowel /u/ at time instant $t \approx 6.4$ ms without radiation ($p = 0$ at the mouth surface). (b) and (d) represent the same results than in (a) and (c), but translating the acoustic pressure level as a displacement over a cut of the vocal tract.

To conclude, the above results highlight the importance of accurate modelling radiation in voice production, because small changes in the reactance and resistance may lead to strong variations in formant location and bandwidth. In this sense, very accurate impedance load models should be used or, otherwise, free-field radiation should be allowed. However, in time domain simulations, impedance loads are difficult to implement, so a good alternative is that of free-field radiation. The time domain FEM approach with PML used in this work could be viewed as a good solution.

5.3 End-corrected vocal tracts

5.3.1 Problem statement

In the previous Section 5.2, it has been shown that poor results are obtained if the computational domain ends at the mouth aperture and an open-end boundary condition $p = 0$ is imposed on it. However, this situation is of particular interest because it allows dealing with small computational domains and thus saving a considerable amount of computational time. Moreover, it bypasses implementing impedance load models, which is a difficult task for time domain simulations. Consequently, these simplifications are often used in literature (e.g., Hannukainen *et al.*, 2007; Švancara and Horáček, 2006; Vampola *et al.*, 2008b). It will be next shown that some of the problems associated to these simplifications can be easily overcome by lengthening the vocal tract through appropriate end correction, at almost no computational cost.

As explained in Section 5.2, the effects of the end correction are mainly associated with reactance so that lengthening modifications will basically produce formant shifts and associated slight amplitude variations. Therefore, no improvements can be achieved on effects due to resistance such as bandwidth increasing. However, in some applications such as the medical ones, formant location turns to be of primary importance to analyze how surgery could modify them (e.g., Švancara and Horáček, 2006), whereas bandwidth accuracy plays a secondary role. Vowel identification (which is typically based on the identification of the first two formants) is another application that could benefit from

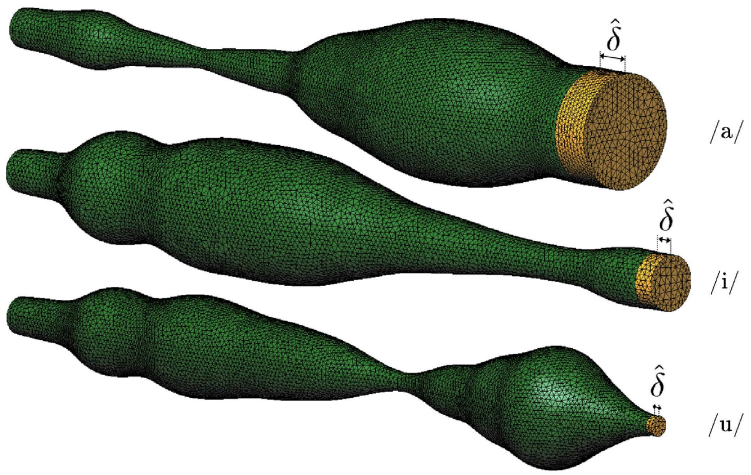


Fig. 5.8: Surface mesh details for vowels /a/, /i/ and /u/ in the simulations with lengthened vocal tracts.

vocal tract lengthening (e.g., Al Bawab *et al.*, 2010). The first apparent difficulty that arises when trying to lengthen the vocal tract through the inertial end correction δ is that the latter is frequency dependent, as it can be seen from Fig. 5.3d. However, this figure also shows that for vowels with small mouth aperture, such as /i/ and /u/, δ is quite constant through all the frequency range, so that an arithmetic average value $\hat{\delta}$ could work as a good approximation.

5.3.2 Lengthening based on the piston set in a sphere model

Let us first take $\hat{\delta}$ as the frequency averaged end corrections of the piston set in a spherical baffle for vowels /a/, /i/ and /u/, the equivalent load model for the free-field radiation case in Section 5.2, in the frequency range from 0 to 5 kHz. The vocal tracts are then lengthened according to their respective $\hat{\delta}$ (see Fig. 5.8). This only results in a slight increase of the computational domain that is still endorsed with an open end boundary condition $p = 0$ at the mouth exit. However, the effects of this correction on the vocal tract transfer functions $H(f)$ will be notorious.

The vocal tract transfer functions $H(f)$ are presented in Fig. 5.9. For each vowel, the free-field radiation, non-lengthened vocal tract (also referred as open-end in some figure legends) and the lengthened cases are compared. We can observe a clear improvement in the formant locations and amplitudes when using the end corrected vocal tracts. Note however, that the formant bandwidths are not corrected given that lengthening cannot account for resistance effects. For a better quantification of this improvement, the

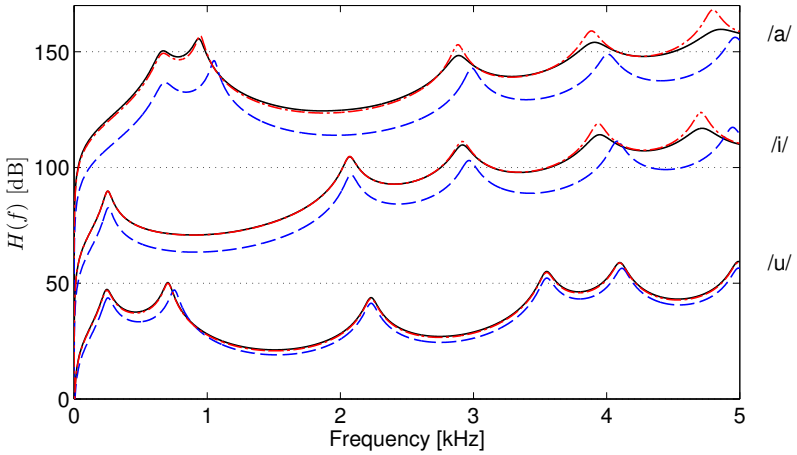


Fig. 5.9: Transfer function $H(f)+50$ dB for vowel /a/, $H(f)$ for vowel /i/ and $H(f)-50$ dB for vowel /u/. (— Free-field radiation, - - open-end, –· lengthened).

differences between the lengthened and non-lengthened cases with respect to the free-field radiation situation are given in Table 5.2. The results of the lengthened case for vowel /u/ are striking but some differences with the free-field radiation case still remain for vowels /i/ and /a/, specially for the high frequency range. This is because of the frequency average of the end correction, which constitutes a rough approximation in the case of large mouth exits (see Fig. 5.10).

Table 5.2: Deviation of formant locations for vowels /a/, /i/ and /u/, for the lengthened vocal tracts with respect to the free-field radiation case (free-field). For comparison, the values for the open-end case (non-lengthened) have been also included.

Vowel	Case	Deviation (free-field)			
		F_1	F_2	F_3	F_4
/a/	lengthened	2	11	-6	-23
	non-lengthened	13	116	99	100
/i/	lengthened	1	0	-1	-11
	non-lengthened	5	9	50	132
/u/	lengthened	0	0	0	0
	non-lengthened	8	48	0	1

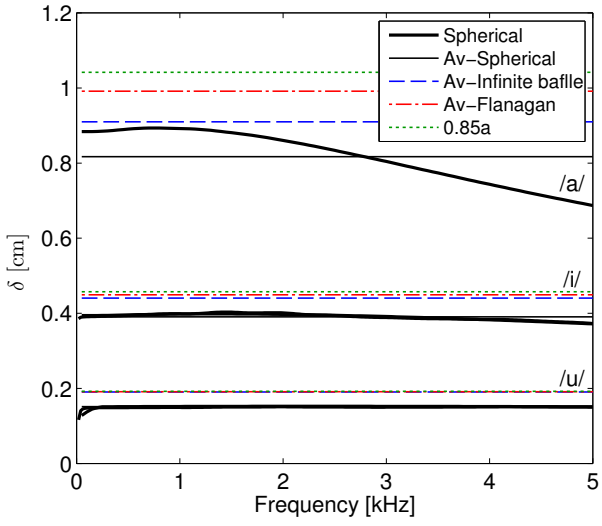


Fig. 5.10: Inertial end correction of the piston set in a spherical baffle, and averaged values computed for the spherical model, the infinite baffle model and the Flanagan and $0.85a$ approximations.

5.3.3 Lengthening based on other radiation models

Finally, to illustrate the benefits of using the spherical model and to highlight the importance of well-modelling the acoustic radiation, the vocal tract has also been lengthened according to other well-known theoretical models. We have used the celebrated end correction $0.85a$, the Flanagan's approximation and the piston in a infinite baffle. The $0.85a$ case is actually an approximation to Flanagan's model, which in turn is an approximation to the piston in an infinite baffle. The first two models assume $ka << 1$, which is not valid for large mouth apertures, as commented in Section 5.2.2.

The frequency averaged inertial end corrections $\hat{\delta}$ have been computed for every model (see Fig. 5.10) and lengthened vocal tracts have also been built according to them. Again, the vocal tract transfer functions $H(f)$ have been calculated and deviations with respect to the free-radiation case have been summarized in Table 5.3. As it could be expected, the errors are progressive, the worst ones corresponding to the $0.85a$ model, followed by Flanagan, the infinite baffle and the spherical one (the prefix "Av" has been added to the legends in tables and figures to remark that the various model end corrections have been frequency averaged). The exactitude of the various models can be also appreciated in Fig. 5.11, where the transfer function for vowel /a/ has been computed according to each model end correction.

Having a look at Table 5.3, it is apparent that bigger errors are obtained for large mouth apertures and for high frequencies (see also Fig. 5.11). The former is logical if one considers the effects of frequency averaging for the inertial end correction (see Fig. 5.10). The

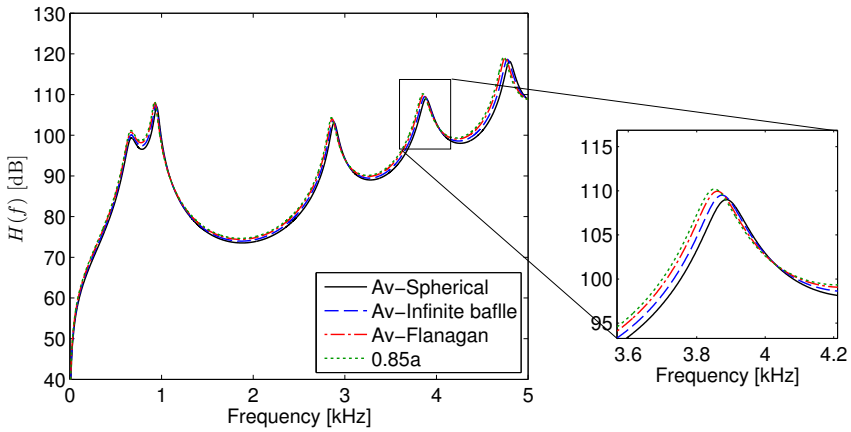


Fig. 5.11: Vowel /a/ transfer functions for lengthened vocal tracts using different approximations for the inertial end correction

Table 5.3: Deviation of formant locations using lengthened vocal tracts with different approximations for the inertial end correction. The obtained values are referenced to the free-field radiation case (free-field).

Vowel	Case	Deviation (free-field)			
		F_1	F_2	F_3	F_4
/a/	Av-Spherical	2	11	-6	-23
	Av-Infinite baffle	0	1	-17	-36
	Av-Flanagan	-2	-7	-27	-50
	$0.85a$	-3	-12	-33	-58
/i/	Av-Spherical	1	0	-1	-11
	Av-Infinite baffle	0	-2	-9	-34
	Av-Flanagan	0	-2	-11	-39
	$0.85a$	0	-2	-12	-42
/u/	Av-Spherical	0	0	0	0
	Av-Infinite baffle	-2	-10	0	0
	Av-Flanagan	-2	-10	0	0
	$0.85a$	-2	-10	0	0

latter could be also expected, given that any error in the average end correction will be more evident for high frequencies (the reactance will be larger and so will be the formant frequency shifts). However, as explained in Section 5.2.4 this cannot be viewed as a general rule due to formant-cavity affiliation. Small errors in the reactance can result in large errors in the formant locations, as happens for vowel /u/ (see Table 5.3). As vowel /u/ has the smallest mouth aperture, its formants should experience almost no deviation for the lengthened vocal tracts. This is in fact the case for the first, third and fourth formants but not for the second one, which corresponds to a front-cavity formant and it is thus strongly influenced by radiation effects. On the other hand, it can be observed in Table 5.3 that for the first two formants of vowel /a/, better results are obtained using the Av-Infinite baffle end correction than the Av-Spherical one. This is so because the Av-Infinite baffle end correction value is closer to the spherical end correction than the Av-Spherical one, for the 0 to 2 kHz frequency range (see Fig. 5.10). Actually, it is to be noted that depending on the type of application one could compute the end correction by frequency averaging only in a limited frequency range (for instance, from 0 to 2 kHz would suffice for vowel identification, whereas the typical range for medical applications comprises from 0 to 5 kHz, see eg., Švancara and Horáček, 2006). Also note that weighted frequency averages could be also used instead of the arithmetic one for computing the end correction, to try to achieve better results.

5.4 Conclusions

In this chapter, the importance of well modelling radiation losses has been illustrated by analyzing their effects on three-dimensional vocal tracts with circular cross-sections set in a spherical baffle. A time domain finite element approach (3D-FEM) combined with a perfectly matched layer to account for free-field radiation has been used.

The radiation impedance for vowels /a/, /i/ and /u/ has been computed, showing a great accuracy when compared to the theoretical model of a piston set in a spherical baffle. Then, the effects of resistance and reactance on vowel vocal tract transfer functions have been analyzed comparing a free-field radiation case versus an open-end case. A bandwidth increase together with a formant location shift to lower frequencies and slight amplitude variations have been observed. These radiation effects are not important for the first formant of vowels, but in contrast to tubes with constant cross-section, they present a rather intricate dependence with resistance and reactance for the remaining formants due to the formant cavity affiliation phenomenon. It turns out that small errors in radiation modelling can lead to strong variations in the vocal tract transfer functions, thus reinforcing the importance of well-modelling this loss mechanism. In terms of vocal tract transfer functions, the used time domain 3D-FEM with free-field radiation has shown better performance than standard 1D models with an impedance load, given that the latter can not deal with non-planar wave propagation within the vocal tract. On the other hand, a non-standard method for detecting front and back cavity formants has been derived. This method only requires two wide-broadband simulations, with and without radiation, and can be viewed as an alternative to the approach where each one of the formant pressure distribution patterns are visually examined one by one (one numerical simulation for each formant) to determine the formant-cavity affiliation. Lastly, it has been shown that for computational domains ending at the mouth aperture, reactance effects can be almost considered by simply lengthening the vocal tract geometry through appropriate end-correction. The additional computational cost of such a procedure is almost negligible. Although probably not useful for the generation of good quality synthetic voice, lengthening could be of value in medical and vowel identification applications.

Finally, we would like to point out that approximating the human head geometry by a sphere is a rough option for a real speaker. In the forthcoming Chapter 6, radiation effects considering more realistic facial surfaces will be examined and contrasted to the results obtained with the spherical approximation.

Chapter 6

Acoustics effects of human head geometry simplifications

One of the key effects to model in voice production is that of acoustic radiation of sound waves emanating from the mouth. The use of three-dimensional numerical simulations allows to naturally account for it, as well as to consider all geometrical head details, by extending the computational domain out of the vocal tract. Despite this advantage, many approximations to the head geometry are often performed for simplicity and impedance load models are still used as well to reduce the computational cost. In this chapter, the impact of some of these simplifications on radiation effects is examined for vowel production in the frequency range $0 - 10$ kHz, by means of comparison with radiation from a realistic head. As a result, recommendations are given on their validity depending on whether high frequency energy (above 5 kHz) should be taken into account or not.

The contents of this chapter have been respectively published and presented in

- Marc Arnela, Oriol Guasch and Francesc Alías (2013), "Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations," Journal of the Acoustical Society of America, 134(4), pp. 2946–2954.
- Marc Arnela and Oriol Guasch (2014), "Validation of the piston set in a sphere model for vowel sound radiation losses against realistic head geometry using time-domain finite-element simulations," 9th International Conference on Voice Physiology and Biomechanics (ICVPB), April 10–12, Salt Lake City, USA.

6.1 Introduction

The importance of accurate modelling of the radiation loss mechanism has been highlighted in the previous Chapter 5. Several strategies to deal with it have been reviewed and compared, which include open-end solutions, impedance load models and free-field approaches. Impedance load models are probably the most extended solution. Their key feature and advantage is that they end the computational domain at the vocal tract exit (mouth) and impose a frequency dependent radiation impedance on it (e.g., the circular piston set in a sphere or in a flat baffle). However, these models make some strong assumptions and can only be seen as simple approximations of the voice radiation mechanism. Clearly, the human head geometry is not that of a flat plane or a sphere; the influence of some head details such as the nose or the lips is not considered and may be of importance, the mouth aperture is clearly not circular and so on. It is the main purpose of this chapter to examine the differences in radiation effects when performing head geometry simplifications, and to analyze the extent to which they can be carried out whilst still fitting the radiation behavior of a realistic head. This has been done by resorting to time domain finite element simulations.

In three-dimensional (3D) numerical computations for voice generation impedance load models have been used to emulate radiation (e.g., Zhou *et al.*, 2008; Vampola *et al.*, 2008a,b), though it seems more natural to consider free-field options allowing for radiation (e.g., Švancara and Horáček, 2006; Takemoto *et al.*, 2010; Vampola *et al.*, 2011; Arnela and Guasch, 2013). The latter directly accounts for radiation effects, though at a higher computational cost. Despite this advantage, the human head geometry has been typically approximated, for instance, as a spherical baffle for simplicity (see e.g., Švancara and Horáček, 2006; Vampola *et al.*, 2011; Arnela and Guasch, 2014b), as done also in Chapter 5 for comparison purposes. We will next show the effects of considering more realistic human heads.

The effects of head geometry simplifications will be analyzed for vowels /a/ and /u/, since they respectively correspond to the limiting cases of large and small mouth aperture. Four different head geometries will be considered: (a) a realistic head geometry with an elliptical cross-section vocal tract that will be used as a reference, (b) a spherical baffle with lips and elliptical vocal tract, (c) a spherical baffle without lips but still with an elliptical vocal tract, and (d) a spherical baffle without lips and with a circular cross-section vocal tract. The influence of these simplifications on the radiation of vowels /a/ and /u/ will be quantified by means of the radiation impedance and its effects on vocal tract transfer functions. The latter will be computed from the numerical solution of the acoustic wave equation, using a time domain finite element method (FEM) with a PML formulation. To

obtain the radiation impedance, use has been made of an adaptation of a two-microphone transfer function method (TMTF) to the numerical framework (see Chapter 4).

The analysis of the results will be split into a low frequency range below 5 kHz, where plane wave propagation dominates and determines voice intelligibility, and a high frequency range up to 10 kHz, whose energy contributes to the perception of voice quality (Monson *et al.*, 2011). Recommendations for head geometry simplifications will be given depending on whether an analysis is to be performed up to 5 kHz or up to 10 kHz, bearing in mind possible perceptual consequences and computational costs.

This chapter is structured as follows. The methodology followed to examine the radiation effects of the head simplifications is presented in Section 6.2. The obtained results are detailed in Section 6.3, where the impact of the considered simplifications on vowel production is discussed. Conclusions close the chapter in Section 6.4.

6.2 Methodology

6.2.1 Human head and vocal tract models

A realistic head model of a male from the 3D modelling company TurboSquid (Bitmapworld, 2005) (see Fig. 6.1) has been chosen. The model is completely articulated with many phonemes for speech, and by means of morph targets (blend shapes) and linear combinations, new phonemes can be generated. In this work, the TurboSquid model has been adapted for vowel production simulations. First, the model has been scaled to have similar dimensions to those of a real male head (see e.g., Burkhard and Sachs, 1975), obtaining 15.6 cm and 20 cm for the head breath and length. The eyes, eyelashes, nostrils and ear canals have been removed and the resulting gaps have been covered with simple surfaces. Next, to simplify the generation of the finite element mesh, the skin has been smoothed to suppress some fine details as wrinkles.

As said in the Introduction, vowels /a/ and /u/ have been selected to examine the radiation effects, since they correspond to the highest and lowest mouth aperture configurations (mouth sections of 4.72 and 0.16 cm^2, respectively). Simplified geometries for their corresponding vocal tracts have been constructed using the cross-sectional areas provided by Story (2008) with a vocal tract length of 16.90 cm for /a/ and 19.59 cm for /u/. The eccentricities of the elliptical mouth apertures in Fromkin (1964) have been applied to them. The vocal tracts have been bent artificially to fit within the 3D head.

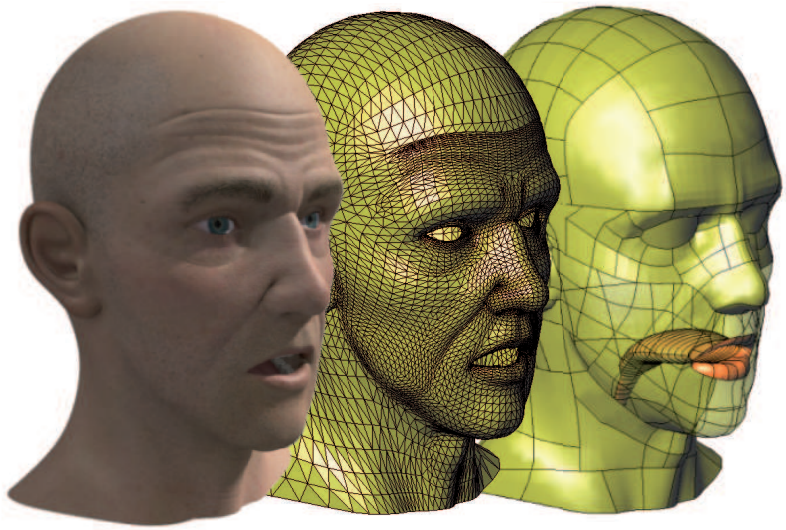


Fig. 6.1: Realistic head geometry for vowel /a/. Left: rendered original head with all details included. Center: head geometry with eyes, nostrils and ear canals removed. Right: final head geometry with inserted vocal tract.

To do so, their centerlines have been smoothly bent forming an angle of 90° with vertex close to their center point, and the cross-sectional areas have been located in the normal direction to the bent centerlines.

The corresponding vocal tract of each vowel has been inserted into the head. For each case, the width and height of the lip opening (see e.g., Fromkin, 1964) in the realistic head have been adjusted to fit the vocal tract exit cross-section. Next, following Story *et al.* (1996), the front termination plane of the vocal tract (also termed the mouth termination plane) has been defined using the technique of Mermelstein (1973). However, it has been observed that this plane can be located between the corner of the mouth and the outer-most points of the lips. Therefore, a complete outline of the vocal tract cross-section cannot be determined in the front termination plane (see Fig. 6.2). The outer-most plane where this can be done is that containing the mouth corner. Yet, the outline at this plane will be not elliptical as those at the exit of the inserted vocal tracts. It is then necessary to place the vocal tract at a certain (small) distance in front of the mouth corner plane, so that its elliptical exit cross-section can be coupled with the non-elliptical outline at the mouth corner plane (the plane of the elliptical exit cross-section of the vocal tract will be hereafter referred to as the coupling plane). This procedure introduces some artificial lengthening; the distance between the coupling plane and the mouth termination plane has to be removed by eliminating the corresponding back end cross-sections of the original

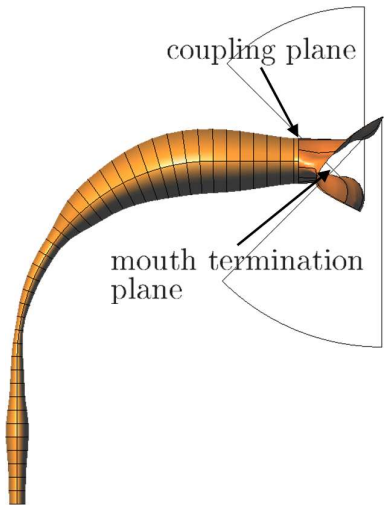


Fig. 6.2: Vocal tract geometry for vowel /a/ coupled with the lips. It is signaled the mouth termination and coupling planes for this case.

vocal tract. The original length of the vocal tract is preserved but now comprises from the glottis to the front termination plane. In Fig. 6.2 it is shown a detail of the coupling between the lips and the vocal tract geometry for vowel /a/, including also the locations of the mouth termination and coupling planes for this case. The above procedure is mandatory for vowel /a/ but much easier for vowel /u/, because the front termination plane already presents a complete almost elliptical outline and can be approximated as being so.

6.2.2 Human head simplifications

To evaluate the impact of geometry simplifications on radiation four cases of decreasing accuracy have been considered (see Fig. 6.3), namely, (a) the realistic head, (b) the spherical head with lips, (c) the spherical head with elliptical mouth aperture, and (d) the spherical head with circular mouth aperture. The first one constitutes the reference model since it contains most geometry details of the head. In the second one, the head has been replaced by a sphere of identical volume (obtaining a radius of $r = 0.1048$ m) though the lips have been kept in the model. The third case corresponds to the second one but with the lips being removed from the spherical baffle. Finally, in the fourth case, circular cross-sections are considered for the vocal tracts instead of elliptical ones. Note that this is the only case with analytical models available for radiation impedance computation. Considering this case is of special interest, despite of its simplicity, because it is widely

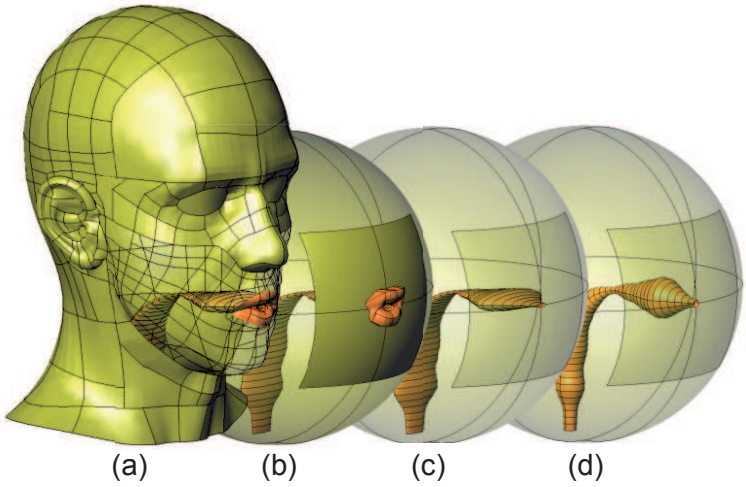


Fig. 6.3: Geometries for vowel /u/ corresponding to the four simplifications described in text: (a) realistic head, (b) spherical head with lips, (c) spherical head with elliptical mouth aperture and (d) spherical head with circular mouth aperture.

used to emulate radiation in voice production both in 1D and 3D simulations. For the cases without lips (c) and (d), the front termination plane of the vocal tract is defined where the major semi-axis (or radius) of the last vocal tract cross-section intersects with the spherical baffle.

In addition to the above cases, an extra test has been carried out to check if the computational domain can be reduced to diminish the computational cost without losing significative quality. Instead of the whole sphere, only a small portion of it surrounding the mouth aperture has been taken into account, thus considerably limiting the outer computational domain (see the shaded region of the spherical baffle in Fig. 6.3).

6.2.3 The wave equation for vocal tract acoustics

The mathematical problem to be solved by the finite element method will be next exposed. Let us denote by Ω the computational domain where acoustic waves propagate. Ω comprises from the vocal tract to the outer region surrounding the head, where waves emanating from the mouth propagate in free field conditions. The boundary of Ω, $\partial\Omega$, can be split into four non-intersecting regions (see Fig. 6.4): Γ_{G} stands for the vocal tract cross-section at the vocal folds position, Γ_{W} corresponds to the vocal tract walls, Γ_{H} to the head contour and Γ_{∞} is a fictitious non-reflecting boundary.

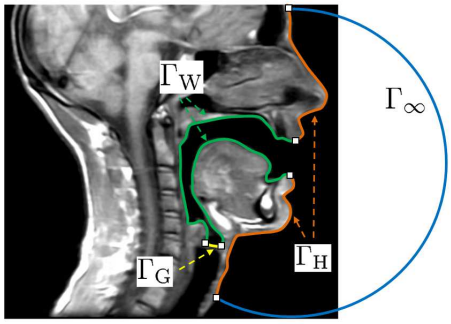


Fig. 6.4: A sketch of the computational domain Ω of Eq. (6.1) in text. Γ_G represents the glottal cross-sectional area, Γ_W the vocal tract walls, Γ_H the human head and Γ_∞ a non-reflecting boundary.

The acoustic pressure due to vowel generation can be obtained from the solution of the wave equation

$$\left(\partial_{tt}^2 - c_0^2\nabla^2\right) p = 0 \qquad \text{in } \Omega,\ t > 0, \tag{6.1a}$$

with boundary and initial conditions

$$\nabla p \cdot \boldsymbol{n} = g \qquad \text{on } \Gamma_{\mathrm{G}},\ t > 0, \tag{6.1b}$$

$$\nabla p \cdot \boldsymbol{n} = -\mu/c_0 \partial_t p \qquad \text{on } \Gamma_{\mathrm{W}},\ t > 0, \tag{6.1c}$$

$$\nabla p \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\mathrm{H}},\ t > 0, \tag{6.1d}$$

$$\nabla p \cdot \boldsymbol{n} = 1/c_0 \partial_t p \qquad \text{on } \Gamma_{\infty},\ t > 0, \tag{6.1e}$$

$$p = 0,\ \partial_t p = 0 \qquad \text{in } \Omega,\ t = 0. \tag{6.1f}$$

In (6.1) c_0 stands for the speed of sound, $p(\boldsymbol{x}, t)$ is the acoustic pressure, $g(t)$ is a function related to the airflow generated by the vocal folds (glottal pulses), $\mu(\boldsymbol{x})$ is a friction coefficient for the losses at the vocal tract walls, $\boldsymbol{n}$ is the normal vector pointing outwards $\partial\Omega$ and $\partial_t \equiv \partial/\partial t$ denotes the partial time derivative. Explicit dependence of variables on space $\boldsymbol{x}$, time t, and angular frequency ω, will be often omitted to shorten notation, except where it could facilitate comprehension.

The boundary condition in Eq. (6.1b) accounts for the volume velocity $Q_g(t) = u_g A_g$ generated by the vocal folds (u_g stands for the inflow velocity and A_g for the area of the boundary Γ_{G}). The time derivative of the volume airflow $\partial_t Q_g$ is provided by most parametric glottal models such as the Rosenberg (Rosenberg, 1971) or LF (Fant *et al.*,

1985) ones. Then, it follows from the momentum equation that

$$\nabla p \cdot \boldsymbol{n} = -\rho_0 \partial_t u_g = -\rho_0 \partial_t Q_g / A_g \equiv g, \tag{6.2}$$

where ρ_0 designates the air density and function $g(t)$ in (6.1b) is defined in the last equality.

The subsequent boundary condition in Eq. (6.1c) considers losses at the vocal tract walls (locally reacting soft walls). It arises from the following reasoning. The impedance of the wall $Z(\boldsymbol{x}, \omega)$ can be used to relate the acoustic velocity at the wall $U(\boldsymbol{x}, \omega)$ to the acoustic pressure $P(\boldsymbol{x}, \omega)$,

$$U(\boldsymbol{x}, \omega) = \frac{P(\boldsymbol{x}, \omega)}{Z(\boldsymbol{x}, \omega)}. \tag{6.3}$$

The momentum equation in the frequency domain can then be written as ($\partial t \leftrightarrow j\omega$, with $j = \sqrt{-1}$)

$$\nabla P(\boldsymbol{x}, \omega) \cdot \boldsymbol{n} = -j\omega\rho_0 \frac{P(\boldsymbol{x}, \omega)}{Z(\boldsymbol{x}, \omega)}. \tag{6.4}$$

Assuming a frequency independent impedance, introducing the boundary admittance coefficient

$$\mu = \frac{\rho_0 c_0}{Z}, \tag{6.5}$$

replacing it in Eq. (6.4) and transforming the resulting expression back to the time domain, results in Eq. (6.1c). At this point, it is convenient to comment that the impedance of the vocal tract wall is not actually frequency independent (see e.g., Ishizaka *et al.*, 1975). However, implementing frequency dependent boundary conditions in the time domain is not a straightforward task (see Nieuwenhof and Coyette (2001) for some steps towards this goal), so that the constant frequency assumption is usually adopted (see e.g., Takemoto *et al.*, 2010). On the other hand, the homogeneous Neumann boundary condition in Eq. (6.1d) simply states that the surface of the head is modeled as a hard wall (i.e., $Z(\boldsymbol{x}, \omega) = \infty$).

Finally, Eq. (6.1e) is the well-known Sommerfeld radiation condition, which guarantees that emitted waves propagate outwards to infinity. However, this condition is only optimal for waves impinging orthogonal onto Γ_∞. To avoid this problem and to perform simulations in a computational domain of reasonable size as well, Eq. (6.1e) is replaced with a perfectly matched layer. PMLs were originally introduced by Berenger (1994) and are designed to absorb waves incident from any direction. The PML formulation of Grote and Sim (2010) has been adapted for our custom code, which solves Eq. (6.1) with the PML using the finite element method (see also Chapter 2 for details on the implemented numerical scheme).

6.2.4 Simulation details

The following computational features have been used to compute the acoustic pressure of vowels. The values $c_0 = 350$ m/s and $\rho_0 = 1.14$ kg/m^3 have been chosen for the speed of sound and air density in Eq. (6.1). In what concerns the boundary conditions, the Gaussian pulse $gp(n)$ (Takemoto *et al.*, 2010) has been used for the volume velocity Q_g in (6.2),

$$gp(n) = e^{[(\Delta t \ n - T_{gp})0.29T_{gp}]^2} [\mathrm{m}^3/\mathrm{s}], \tag{6.6}$$

with $T_{gp} = 0.646/f_0$ and $f_0 = 10$ kHz. This pulse has been low pass filtered with a cutoff frequency of 10 kHz, to avoid the appearance of high frequency numerical errors above the maximum frequency of analysis (10 kHz). In (6.1c), a boundary admittance coefficient of $\mu = 0.005$ has been chosen to account for wall losses. This value corresponds to the impedance of the vocal tract tissue $Z_w = 83666$ kg/m^{2}s (see Švancara and Horáček, 2006).

With regard to the computational domain, the vocal tract and head geometries described in previous Sections have been employed. The head has been surrounded with a volume of dimensions $0.24 \times 0.28 \times 0.28$ m to allow sound waves emanate from the mouth and, in turn, this volume has been surrounded with a perfectly matched layer, 0.1 m width, to absorb any incident wave. The PML has been configured to get a reflection coefficient of $r_\infty = 10^{-4}$. The lower part of the neck has been immersed within the PML as in Mokhtari

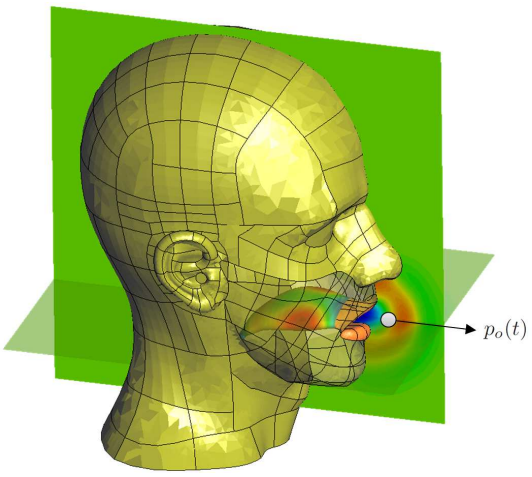


Fig. 6.5: Wave propagation inside the vocal tract and emanating from mouth aperture at time instant $t = 0.75$ ms for vowel /a/ with a realistic head. It represents all the volume for the vocal tract, while for the outer volume a cut in the midsagittal plane is done for visualization purposes. The PML region is not shown. The white dot denotes the position where the acoustic pressure $p_o(t)$ is captured.

et al. (2008), to avoid the sound waves following artificial paths through the underside of the head model. On the other hand, when only a small portion of the spherical baffle is considered (additional test case), the outer volume dimensions become $0.09 \times 0.1 \times 0.1$ m. The computational domains have been meshed using tetrahedral elements. The mesh size h ranges from $h = 0.001$ m within the vocal tract to $h = 0.003$ m in the mouth and nose immediate outer region, $h = 0.005$ m in the most outer volume, and $h = 0.0075$ m in the PML. This results in, approximatively, 3.5 million and 1 million elements for the complete and the reduced computational domains, respectively.

Concerning the time evolution, a sampling rate of $f_s = 1/\Delta t = 2000$ kHz has been used and the total duration of the simulation has been of $T = 25$ ms. Fig. 6.5 shows a snapshot at time instant $t = 0.75$ ms of the acoustic pressure field for vowel /a/, with a realistic head. The influence of high order modes can be identified within the vocal tract and it can be appreciated how spherical front waves emanate from the mouth propagating outwards.

6.2.5 Transfer functions and radiation impedances

The time evolution of the FEM computed acoustic pressure $p_o(t)$ has been collected at a distance of 2 cm from the mouth termination plane (see white dot in Fig. 6.5). From its Fourier Transform, the vocal tract transfer function (VTTF) can be obtained as

$$H(f) = \frac{P_o(f)}{Q_g(f)}. \tag{6.7}$$

In order to compute the radiation impedance Z_r, the procedure in Arnela and Guasch (2013) has been followed, which consists in adapting a TMTF method to the numerical framework (see also Chapter 4). In brief, the vocal tract geometry is replaced by a lossy impedance duct of length 0.1 m, having the same shape and cross-sectional area than the vocal tract exit. An admittance coefficient value of $\mu = 0.008$ is imposed to the impedance duct of vowel /a/, in the realistic head model. This artificial value has been deemed appropriate from numerical experiments, showing a good compromise between the computational cost and the accuracy of the results. In order to get the same absorption for the remaining impedance ducts, their boundary admittances have been computed imposing their complex axial wavenumbers to equal that of the /a/ impedance duct in the realistic head model, which yields

$$\mu_m = \mu_n \frac{\mathcal{P}_n S_m}{\mathcal{P}_m S_n}. \tag{6.8}$$

In Eq. (6.8), subindex n denotes the reference case (/a/ impedance duct in the realistic head) and subindex m the new case for which μ has to be computed. S is the cross-section area of the impedance duct and $\mathcal{P}$ its perimeter. Then, a numerical simulation is

performed and the acoustic pressures $p_1(t)$ and $p_2(t)$ at two nodes inside the impedance duct, separated a distance $s = 1$ cm apart, become tracked. These nodes are chosen to be located in the centerline of the impedance duct. This allows extending almost two times the upper frequency limit of the TMTF method so that the radiation impedance of the planar mode at high frequencies can be computed. From the Fourier Transforms of $p_1(t)$ and $p_2(t)$, $H_{12}(f) = P_2(f)/P_1(f)$ is calculated to get the reflection coefficient,

$$\mathcal{R} = \frac{H_{12} - e^{-jk_z s}}{e^{jk_z s} - H_{12}} \, e^{j2k_z|x_1|}, \tag{6.9}$$

where x_1 is the distance from the impedance duct exit where $p_1(t)$ is collected and k_z is the complex axial wave number. The normalized specific acoustic impedance is computed by means of

$$Z_r = \frac{1 + \mathcal{R}}{1 - \mathcal{R}}. \tag{6.10}$$

Finally Z_r becomes split in terms of its real R_r and imaginary X_r components, $Z_r = R_r + jX_r$. The specific acoustic impedance Z_r' can be recovered from $Z_r' = Z_0 Z_r$, with Z_0 standing for the characteristic impedance of the medium,

$$Z_0 \simeq \rho_0 c_0 \left(1 - j\frac{\mu\mathcal{P}}{k_0 S}\right)^{-1/2}. \tag{6.11}$$

Note that for large ducts with small wall losses Eq. (6.11) yields the usual value $Z_0 \approx \rho_0 c_0$. Another parameter that will be considered in forthcoming Sections is the inertial end correction δ, which is related to the reflection coefficient through (e.g., Selamet *et al.*, 2001)

$$\mathcal{R} = |\mathcal{R}|e^{j\theta} = -|\mathcal{R}|e^{-j2k\delta}, \tag{6.12}$$

with $|\mathcal{R}|$ being the modulus of $\mathcal{R}$ and θ its argument.

6.2.6 Acoustic pressure distribution of formants

For a better comprehension of some phenomena like the formant cavity affiliation (see Fant, 1970), the acoustic pressure distribution within the vocal tract for a given formant has been also computed. To do so, the same configuration described in Section 6.2.4 has been used, but a sinusoidal signal has been imposed at the glottis cross-sectional area instead of a gaussian pulse. The sinusoid has the frequency of the formant that it is to be analyzed. After some time steps (~ 1.5 ms), the acoustic pressure within the vocal tract becomes stationary, and the pressure pattern of the formant can be observed.

6.3 Results and discussion

6.3.1 Analysis of the head simplifications

Fig. 6.6 presents the computed radiation impedance Z_r split in terms of resistance R_r and reactance X_r, the inertial end correction δ and the vocal tract transfer function $H(f)$, for vowels /u/ (first row in figure) and /a/ (second row in figure), and for the different head geometry configurations: (a) realistic head, (b) spherical baffle with lips, (c) spherical baffle with elliptical mouth aperture but no lips, and (d) spherical baffle with circular mouth aperture and no lips. For brevity, these configurations will be designated hereafter as the realistic case, the lip case (or the case with lips), the elliptical case and the circular case. For completeness, the first four formant values for each case and their deviations with respect to the realistic case are listed in Table 6.1, and their bandwidths in Table 6.2. In what follows, the differences in their computed radiation impedance will be analyzed, as well as how these differences affect the vocal tract transfer functions (VTTF). Two frequency ranges will be considered, that below 5 kHz corresponding to the plane propagation regime, and that above 5 kHz, where high order modes can be excited.

Effects on the radiation impedance

For the vowel /u/, it can be observed from the impedance values of Fig. 6.6a that simplifying the head geometry as a simple baffle with lips yields close results to those of the most realistic case. Some differences can be observed above $\sim$ 5 kHz, however, for both the resistance R_r and the reactance X_r. In contrast, removing the lips from the spherical baffle (elliptical and circular cases) results in a large decrease of X_r for the whole frequency range, and of R_r for frequencies above $\sim$ 5 kHz. Note that the elliptical and circular configurations present almost the same values of R_r for all frequencies, X_r being slightly smaller for the elliptical case.

Some further insight can be gained by resorting to the corresponding inertial end corrections δ's, plotted in Fig. 6.6b. As one could have expected, the inclusion of lips results in an increase of δ (δ being almost identical for the elliptical and circular cases). These differences can be approximated by a constant offset, which affects the radiation impedance for the following reasoning. On the one hand, a positive offset in δ is translated to X_r as an increase of the slope (see Fig. 6.6a). On the other hand, when $R_r^2 + X_r^2 << 1$, the phase θ of the reflection coefficient will only depend on the reactance X_r, to a first order approximation. Consequently, it follows from (6.12) that the same will hold true

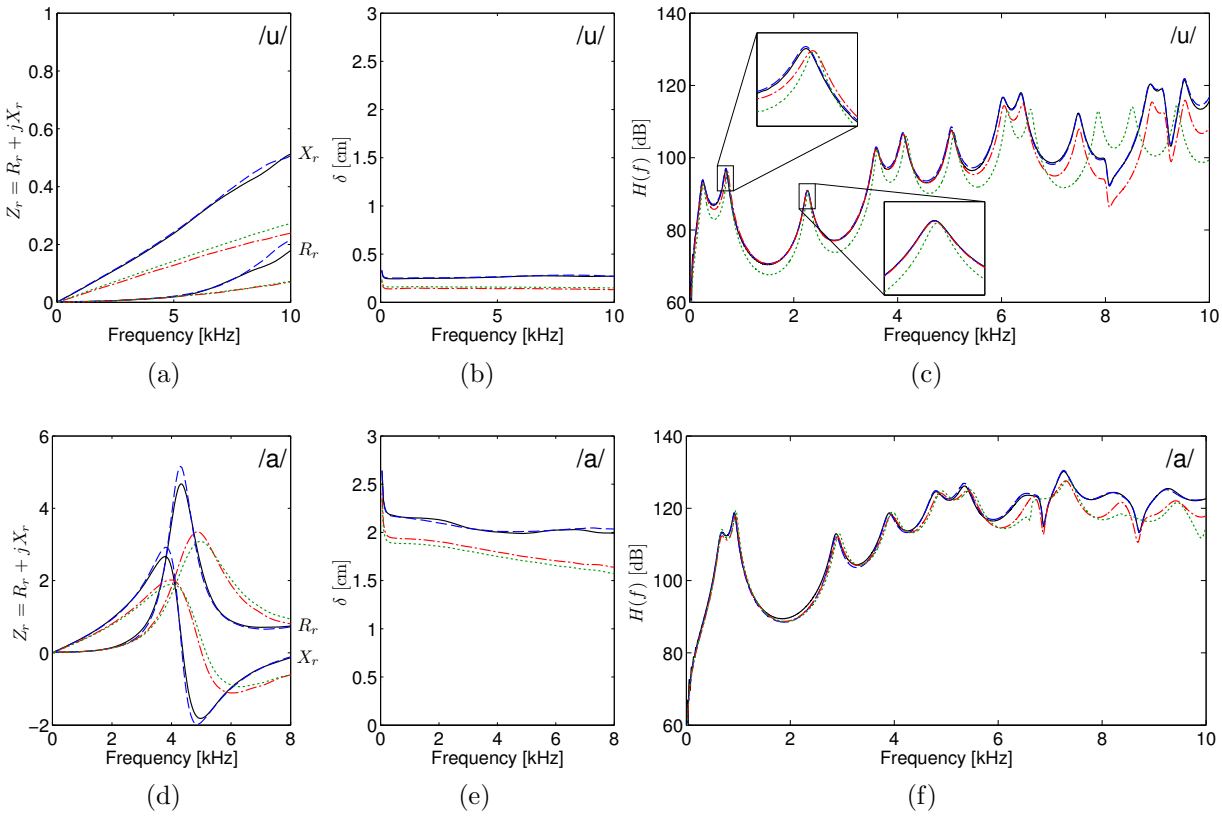


Fig. 6.6: (a) and (d) Radiation resistance R_r and reactance X_r, (b) and (e) inertial end correction δ, and (c) and (f) vocal tract transfer functions $H(f)$ of vowels /a/ and /u/ for different head geometry simplifications. All functions are computed up to 10 kHz except for the radiation impedance and inertial end correction of vowel /a/, which are limited to 8 kHz because of the TMTF method limitations at the onset of higher order modes (see Chapter 4). (— Realistic case, - - Lip case, –· Elliptical case, · · · Circular case).

for δ. This is what happens in the region below $\sim$ 5 kHz, where only the reactance X_r is modified. However, when this condition is no longer satisfied, i.e. above $\sim$ 5 kHz, the resistance R_r also increases.

With regard to vowel /a/, Fig. 6.6d shows the obtained impedance values computed at the coupling plane (see Section 6.2.1). In contrast to vowel /u/, where only plane waves propagate for the whole examined frequency range, the planar wave assumption is only valid up to $\sim$ 4 kHz in the /a/ impedance duct; high order modes also being excited beyond this frequency. However, and as mentioned in Section 6.2.5, it is still possible to compute the radiation impedance of the planar mode up to $\sim$ 8 kHz by appropriate selection of the nodes where the acoustic pressure is collected. Withal, it

is noted that an accurate characterization of the radiation impedance accounting for all involved modes would require the use of a matrix representation, as described in Zorumski (1973). However, most impedance load models commonly used to emulate radiation, e.g., the piston set in a baffle, only consider the planar mode for the whole frequency range (see e.g., Zhou *et al.*, 2008; Vampola *et al.*, 2008a) and could be thus directly compared with the results reported here. In Fig. 6.6d, the computed impedance ressembles the input impedance of a duct with an open end termination because the coupling plane is located inside, though close to the exit, of the vocal tract. Again, more insight can be obtained from the inertial end correction, now shown in Fig. 6.6e. As already observed for vowel /u/, the case with lips presents similar values to those of the realistic case. When the lips are not taken into account, the inertial end correction δ clearly diminishes, being slighter greater for the elliptical case than for the circular one.

Effects on the vocal tract transfer function

The next step consists in analyzing the influence of the observed differences in the radiation impedances of the four configurations, on the VTTFs. The computed transfer functions for each case are plotted in Fig. 6.6c for the vowel /u/. Obviously, a variation in the radiation impedance will not produce a new peak or valley on the VTTF; only changes in formant location, bandwidth and amplitude are expected. However, large variations are observed for the circular case when compared to the other ones beyond the plane wave propagation regime (say above ~ 5 kHz). Recall that elliptical cross-section vocal tracts have been used for the realistic, lips and elliptical cases (see Section 6.2.2) so that their cross-mode formants will clearly differ from those of the circular case.

Focusing on the lower frequency region of the VTTF (below ~ 5 kHz), and analyzing some formants, it can be observed for instance, that the second formant is practically identical for the lip and realistic cases (see zoom in Fig. 6.6c and also Table 6.1). In contrast, when lips are not considered the formant location is shifted up. Besides, comparing the two cases without lips, only small differences in location can be observed. The shifts in the formant location are the effects of reactance X_r variation (see e.g., Kinsler *et al.*, 2000). In general, it is observed that the larger its value the stronger the shift of formants to lower frequencies (see Fig. 6.6a and Fig. 6.6c respectively). However, this rule is not always fulfilled, as happens for example, for the third formant in Fig. 6.6c (see also Table 6.1). Almost no differences can be appreciated among cases although the deviations in reactance values are higher than for the second formant (see Fig. 6.6a). This can be explained by means of the formant-cavity affiliation concept (see e.g., Fant, 1970), where each formant is interpreted as a resonance of a specific vocal tract cavity. For vowel /u/, the first,

third and fourth formants correspond to back-cavities, whereas the second formant is produced by a front-cavity (see Apostol *et al.*, 2004). To illustrate this concept, Fig. 6.7 represents the acoustic pressure field for the second and third formants of vowel /u/ in the realistic case, computed as described in Section 6.2.6 (see also Arnela *et al.*, 2012, where more representations of the formant pressure pattern for vowel /u/ are shown). It can be observed that for the second formant the cavity close to the mouth corresponds to a high acoustic pressure area, whereas for the third formant the highest values are achieved at the cavity close to the glottis. Front-cavity resonances are strongly influenced by acoustic radiation, but the latter's effect on back-cavity resonances is much weaker (Stevens, 2000). This explains why radiation effects are noticeable for the second formant of vowel /u/, whereas no changes can be appreciated for the third formant. In what concerns formant bandwidths (see Table 6.2), no significant differences are observed in the first formants between configurations, with the exception of the circular case which presents lower values (see also e.g. the zoom for the third formant in Fig. 6.6c). These variations are attributed to the vocal tract shape (circular vs elliptical) rather than to radiation effects, since the radiation resistance has a very similar behaviour in the four configurations for this frequency regime (see Fig. 6.6a).

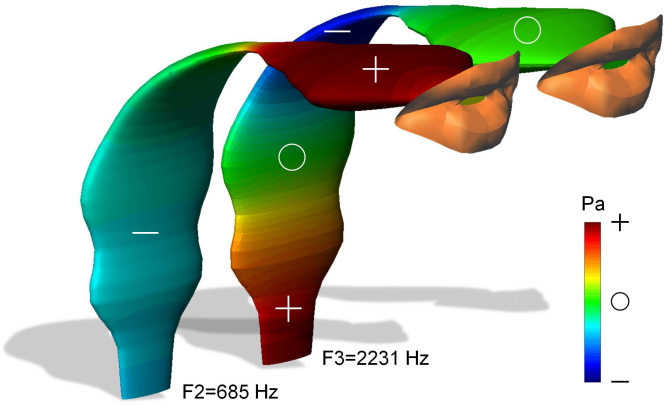


Fig. 6.7: Snapshot of the acoustic pressure for the second (F2) and third (F3) formants of vowel /u/ at time instant $t \approx 7.6$ ms. The regions of the vocal tract where the acoustic pressure is positive, negative and close to zero are represented respectively by crosses, horizontal segments and circles.

Table 6.1: First four formant values in Hz for vowels /a/ and /u/ in the realistic case, the case with lips, the elliptical case and the circular case. Formant deviations ($\Delta F/F$) in % with respect to the realistic case are also indicated.

Vowel	Case	Formant value				Formant deviation			
		F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4
/a/	Realistic	686	924	2884	3911	-	-	-	-
	Lip	686	924	2884	3908	0	0	0	-0.08
	Elliptical	691	945	2909	3946	0.73	2.27	0.87	0.9
	Circular	686	953	2938	3969	0	3.14	1.87	1.48
/u/	Realistic	250	696	2264	3588	-	-	-	-
	Lip	250	695	2264	3588	0	-0.14	0	0
	Elliptical	256	723	2264	3589	2.4	3.88	0	0.03
	Circular	256	729	2275	3612	2.4	4.74	0.49	0.67

Table 6.2: First four formant bandwidths in Hz for vowels /a/ and /u/ in the realistic case, the case with lips, the elliptical case and the circular case.

Vowel	Case	Bandwidth value			
		BW_1	BW_2	BW_3	BW_4
/a/	Realistic	126	91	164	256
	Lip	126	93	154	244
	Elliptical	126	91	171	230
	Circular	102	76	153	237
/u/	Realistic	101	88	116	138
	Lip	100	87	116	138
	Elliptical	100	87	116	137
	Circular	70	64	84	96

With regard to the high frequency range (above ~ 5 kHz), not only variations in formant location become apparent, but also some changes in amplitudes can be observed (see Fig. 6.6c). The circular case will be left apart because, as mentioned, at these frequencies it will not match at all with the configurations having elliptical vocal tracts. A close look at Fig. 6.6c reveals some very small amplitude variations between the lip and the realistic cases, above ~ 8 kHz. In addition, the amplitude of the elliptical case gradually decreases with respect to the realistic case beyond ~ 5 kHz. The variations in amplitude are due to the resistance effects. Above ~ 5 kHz, notable differences are found in the resistance R_r values (see Fig. 6.6a). The resistance for the lip case is higher than for the realistic case, and much higher than for the elliptical one. The resistance is related to the transmitted power (see e.g., Kinsler *et al.*, 2000), so a lower value of resistance will result in a less energetic vocal tract transfer function (see respectively Fig. 6.6a and Fig. 6.6c).

Concerning vowel /a/ (see Fig. 6.6f), similar conclusions can be drawn to those for vowel /u/. In the low frequency region, only small variations in the formant locations are appreciated (see also Table 6.1), and lower bandwidth values are observed for the circular case (see Table 6.2) which are attributed again to the vocal tract shape. Above ~ 5 kHz the circular case clearly distances from the other ones. Shifts in the formant location and differences in the amplitudes of the transfer functions can be observed between the realistic, lip and elliptical cases.

6.3.2 Impact of the geometry simplifications in vowels

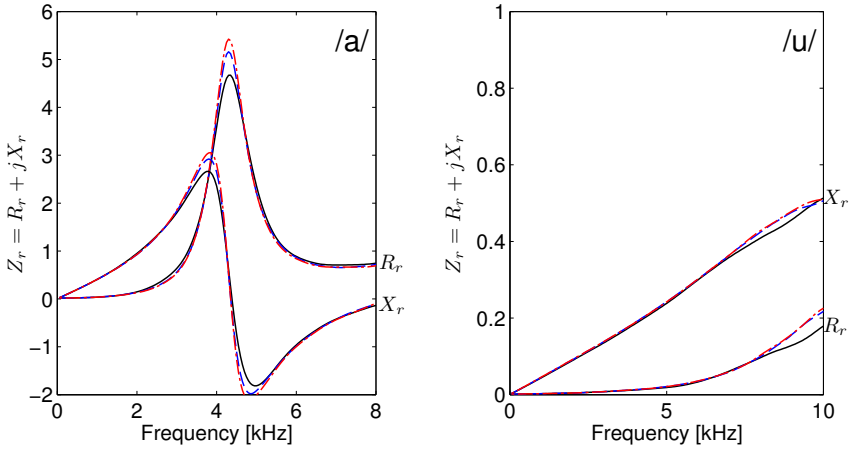
The results presented in Section 6.3.1 indicate that, although large deviations can be found among impedance values, these only result in small differences in the formant locations for the plane wave propagation regime (below ~ 5 kHz). Since, in general, these differences are less than $3-5\%$ (see Table 6.1), one could consider them perceptually not relevant (Flanagan, 2008). Therefore, in this range any of the studied approximations could be used to properly emulate the radiation of vowels. However, if one intends to extend the frequency range above ~ 5 kHz, the lips play a key role as they will induce not only differences in the formant location but also an increase of the transfer function level. The latter ranges from about +1 dB to +6 dB in the frequency range $5-10$ kHz. These level deviations can be considered important because they can modify the naturalness of the generated vowel (see e.g., Monson *et al.*, 2011, where minimum difference limen scores of 1 dB are given for the 8 kHz octave band). Consequently, it is highly recommendable to deal with simulations allowing for net free-field propagation (instead of using impedance load models), as well as to include the lips in the geometry, to correctly reproduce the high frequency content of a vowel. It has also been observed that other head geometry

details such as the nose, head shape, etc., play a secondary role, since their influence on transfer functions can be considered negligible from a perceptual point of view (even though impedance value differences become apparent). They will only be significant for off-axis voice generation (not aligned in front of the mouth), involving three-dimensional full simulation environments where the diffraction of the head can play an important role. In summary, to correctly account for perceptually relevant radiation effects in the generation of a vowel from a realistic head, in the frequency range $0 - 10$ kHz, it suffices to take the case of a spherical baffle that incorporates the lips.

6.3.3 Reduction of the computational domain

It has been mentioned above that it is quite common to use impedance load models at the exit of the vocal tract in some 3D numerical simulations, to account for free-field radiation effects. The advantage of doing so is a great reduction of the computational domain and thus, of the total duration of the simulation. However, in the preceding Section 6.3.2 it has been resolved that including the lips is essential to correctly simulate vowels in the high frequency range, which requires simulating waves emanating from the mouth. Unfortunately, doing so as in the simulations carried out so far in this work, is very time consuming because of the need of including the propagating domain outside the vocal tract. Moreover, a PML is needed to avoid waves reflecting from the computational domain boundaries.

In view of the results discussed in Section 6.3.2, this problem can be partially mitigated by making some further geometry simplifications to reduce the computational cost. Given that the shape of the head does not substantially influence radiation, so that the head can be approximated by a sphere to a good extent, one may wonder what happens if only a portion of the sphere surrounding the lips is taken into account in the simulations. This point has been checked by making simulations that only consider the shaded region of the sphere in Fig. 6.3 and a small outer domain in front of it. The obtained results for the radiation impedance of vowels /a/ and /u/ are shown in Fig. 6.8. This figure also includes the solution for the realistic and lip cases in the complete computational domain, as references. As observed, there is almost no departure from the solution for the lip case, hence, no perceptually relevant changes are to be expected in the corresponding VTTF. It should be noted however, that in practice there is no magic rule to determine to which extent the computational domain can be reduced without affecting the computed results, as many aspects are to be considered. On the one hand, a certain volume for free wave propagation outside the mouth is needed to account for the effects of the head, or sphere, region close to the lips. On the other hand, some issues arise related to the



(a) (b)

Fig. 6.8: Radiation resistance R_r and reactance X_r of vowels /a/ and /u/ when the computational domain is reduced. (— Realistic case, - - Lip case with the complete computational domain, –· Lip case lips with the reduced computational domain).

practical implementation of the PML. To facilitate FEM programming, and depending on the PML formulation, it is recommendable for it to have a simple geometry (cubic o spherical), which imposes some limitations to the reduction process. Moreover, a smooth transition from the very fine mesh inside the vocal tract to the coarse one at the PML is needed to avoid spurious numerical errors. All in all, this means that in practice one has to perform some tests to determine which configuration yields a reasonable balance between the exactitude of the results and the reduction of the computational cost, taking into account all the involved factors.

Using the finally proposed simplification, the number of elements in the finite element mesh was reduced by a factor of $\sim$ 3.5 (see Section 6.2.4 for mesh details), which resulted in a factor $\sim$ 4 decrease of the computational time (numerical simulations were performed in a serial computing system with processor Intel(R) Core(TM)i7 2.67 GHz). The simulations with the whole domain lasted about $\sim$ 36 hours for each analyzed case and $\sim$ 9 hours when the reduced domain was considered. Therefore, the proposed geometry simplification allows a strong saving in computational time and memory at no quality cost in vowel production, from a perceptual point of view.

6.4 Conclusions

The influence on the vowel acoustic radiation of several geometry simplifications of a realistic human head has been studied in this work. The main conclusions derived from the analysis of the results can be summarized as follows. First, the numerical tests indicate that for the examined frequency range ($0 - 10$ kHz) the role played by most head appendices (e.g., the nose), as well as by the head shape, is perceptually irrelevant, with the very important exception of the lips. Second, though large differences can be appreciated in the radiation impedance for the plane wave propagation regime ($0-5$ kHz) among the tested cases, these differences only manifest as small formant location shifts in the vocal tract transfer functions, which are not relevant from a perceptual point of view. Therefore, in this regime any of the analyzed simplified models could be used to properly emulate vowel radiation. Actually and as a consequence, if the goal was to develop a speech synthesizer/simulator that could run in near real-time, the impedance load approach based on a piston set in a spherical baffle could be used instead, as a valid alternative to the by far more computationally expensive free-field radiation simulations. Third and in contrast, for the range $5-10$ kHz, the lips play a significant role and not only produce large deviations in the radiation impedance but also in the vocal tract transfer functions. Therefore, for an accurate modelling beyond 5 kHz the spherical baffle with lips is recommended. Besides, note that it has been shown that it is not necessary to model the whole pressure field around the sphere; a small portion of it suffices which results in a reduction of the computational cost by a factor of ~ 4 without a loss of quality.

Finally, it is worth mentioning that conclusions on the influence of the geometry features of a realistic head on radiation have been based on simplifications from a single reference case. Ideally, one should work with an ensemble of samples for the conclusions to be generalized with statistical confidence. However, in view of the presented results, we do not have the impression that the use of different head geometries could substantially modify the conclusions reached so far. More realistic vocal tracts (e.g. generated from MRI) could also have been used instead of the simple elliptic and circular ones, but again similar conclusions could be expected, given that these resulted from qualitative comparison between models rather than from quantitative values. It would be worthwhile testing in future works the behavior when covering the whole audible frequency range, up to ~ 20 kHz, as well as to analyze radiation effects for other phonemes.

Chapter 7

Two-dimensional simulations with 3D behavior

Two-dimensional (2D) numerical simulations of vocal tract acoustics may provide a good balance between the high quality of three-dimensional (3D) finite element approaches and the low computational cost of one-dimensional (1D) techniques. However, 2D models are usually generated by considering the 2D vocal tract as a midsagittal cut of a 3D version, i.e. using the same radius function, wall impedance, glottal flow and radiation losses as in 3D, which leads to strong discrepancies in the resulting vocal tract transfer functions. In this work, a four step methodology is proposed to match the behavior of 2D simulations with that of 3D vocal tracts with circular cross-sections. First, the 2D vocal tract profile becomes modified to tune the formant locations. Second, the 2D wall impedance is adjusted to fit the formant bandwidths. Third, the 2D glottal flow gets scaled to recover 3D pressure levels. Fourth and last, the 2D radiation model is tuned to match the 3D model following an optimization process. The procedure is tested for vowels /a/, /i/ and /u/ and the obtained results are compared with those of a full 3D simulation, a conventional 2D approach and a 1D chain matrix model.

This chapter is based on the following works

- Marc Arnela and Oriol Guasch (2014), "Two-dimensional vocal tracts with three-dimensional behavior in the numerical generation of vowels," Journal of the Acoustical Society of America, 135(1), pp 369-379.
- Marc Arnela and Oriol Guasch (2014), "Three-dimensional behavior in the numerical generation of vowels using tuned two-dimensional vocal tracts," 7th Forum Acusticum, September 7-12, Krakow, Poland.

7.1 Introduction

In previous chapters of this thesis we have focused on three-dimensional (3D) aspects of the vocal tract acoustics simulation of vowels. Although very accurate results can be achieved, an associated problem to 3D simulations is that of strong computational costs, which limit their use for those applications that aim at working in real-time (e.g., articulatory speech synthesis). Alternatively, to reduce these high computational times and also to diminish the memory requirements, two-dimensional (2D) approaches have been attempted (Speed *et al.*, 2009; Howard *et al.*, 2009; Mullen *et al.*, 2007, 2006; Kako and Touda, 2006), which, in some cases, have been carried out in real-time (see e.g., Mullen *et al.*, 2007). 2D models present some improvements with respect to the 1D types, in terms of the naturalness of the perceived sound (see e.g., Mullen *et al.*, 2006; Howard *et al.*, 2009) or in terms of control flexibility (see e.g., Mullen *et al.*, 2006, where linear control over formant bandwidth is achieved thanks to the second dimension). However, 2D simulations performed to date show large discrepancies when compared to 3D results, in terms of formant locations and bandwidths in the vocal tract transfer functions. It is the main goal of this chapter to suggest a methodology to carry out 2D simulations that match the acoustical behavior of 3D vocal tracts with circular cross-sections.

At this point it is important to note that it is impossible to exactly recover the acoustic performance of a 3D vocal tract with a 2D simulation. The high frequency modes involving the third dimension can not logically be accounted for in the 2D case. In addition, the nature of the wave propagation is different in two and three dimensions. Although a point-like impulse propagates on an infinitely thin wavefront at the speed of sound in 3D free field conditions, this is not the case in two dimensions, where the arrival of a pulse at a field point is followed by a long tail that asymptotically tends to zero. This afterglow phenomenon is characteristic of the wave equation Green function in even dimensions, and some ways to compensate for it in free field, and in 2D simulations of room acoustics as well, have been recently proposed (e.g., Spa *et al.*, 2010; Escolano *et al.*, 2013). Nonetheless, afterglow compensation has not been considered in this work. Though it could have some influence in voice synthesis due to its filtering effect, it will be shown that a close resemblance between 2D and 3D vocal tract transfer functions can already be achieved without taking it into account.

The basic problem with current 2D simulations arises from the fact that 2D vocal tract models are generated as midsagittal cuts of 3D vocal tracts, i.e. using the same radius function, wall losses, glottal flow and radiation model as in 3D. However, four simple steps can be followed to tune a 2D model so that it conforms the acoustic behavior of a circular

3D vocal tract. In the first step, the 2D vocal tract profile is modified so that its formant locations match with those of the 3D case. In the second one, wall losses are allowed to vary in space so as to tune the formant bandwidths. The third step consists in modifying the airflow generated by the vocal folds to recover the 3D pressure levels and, finally, in the fourth step an optimization process is followed to attain a 3D behavior for the 2D radiation impedance. Though simple impedance models could be used for the latter, their implementation in the time domain is by no means straightforward. Alternatively, one can naturally account for radiation by extending the computational domain out of the vocal tract and by making use of infinite elements or perfectly matched layers, to avoid spurious reflections at the domain boundaries. The latter has been adopted in this work and implemented using a custom time-domain FEM code (Arnela and Guasch, 2013).

The four-step 2D vocal tract tuning has been tested for vowels /a/, /i/ and /u/, using the cross-sectional areas provided by Story (2008) for the vocal tract geometries. Although the simulations in this work have been carried out using FEM, any other numerical approach could benefit from the proposed methodology. At this point, it should be noted that in the case of 1D vocal tracts, tuning procedures have been also carried out, though for different purposes than the ones aimed in this work (see e.g., Story *et al.*, 2001; Story, 2006)

The chapter is structured as follows. In Section 7.2, the proposed methodology to construct the tuned 2D vocal tract model is exposed. The results obtained for the three cardinal vowels are discussed in Section 7.3 and contrasted to 3D models, to the standard 2D approaches based on 3D midsagittal cuts, and to a simple 1D frequency domain model, based on standard chain matrices (similar to that of Sondhi and Schroeter, 1987). Conclusions are presented in Section 7.4.

7.2 Methodology

7.2.1 Problem statement

The 3D model to be mimicked with a 2D model is shown in Fig. 7.1. It consists of a 3D vocal tract with circular cross-sections embedded on a sphere of radius a_s, representing the human head (see e.g. Chalker and Mackerras, 1985). The vocal tract is characterized by means of a vocal tract radius function $a_{\mathtt{3D}}(x)$ defined along the vocal tract midline. This has been extracted from the cross-section areas $S(x)$ provided in Story (2008). A boundary admittance $\mu_{\mathtt{3D}}(x)$ is imposed at the vocal tract walls to account for wall losses and a volume velocity $Q_{\mathtt{3D}}(t)$ is introduced at the glottal cross-sectional area, to generate

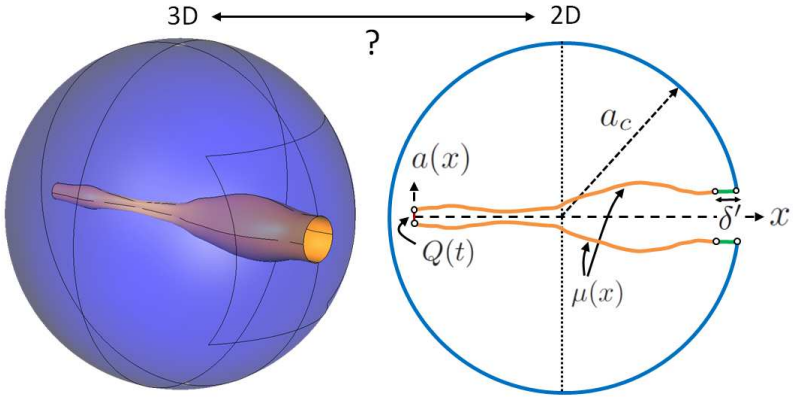


Fig. 7.1: A sketch showing left a 3D vocal tract geometry with circular cross-sections set in a spherical baffle, and right, the equivalent 2D model with the parameters to be tuned: the radius function $a(x)$, the wall boundary admittance $\mu(x)$, the volume velocity at the glottal cross-section $Q(t)$, the radius of the circular baffle a_c and the length δ' of a small duct coupled at the mouth cross-section.

a given sound. Radiation losses are naturally accounted for by allowing sound waves emanate from the mouth aperture and propagate outwards.

The problem to be solved consists in finding a configuration for the 2D model that conforms the behavior of the 3D model described above (see Fig. 7.1). As commented in the Introduction, the standard procedure followed to build a 2D model relies on performing a midsagittal cut of the 3D vocal tract. Actually, the following equivalences are usually made

$$a_{2D}(x) = a_{3D}(x), \tag{7.1a}$$
$$\mu_{2D}(x) = \mu_{3D}(x), \tag{7.1b}$$
$$Q_{2D}(t) = Q_{3D}(t), \tag{7.1c}$$
$$a_c = a_s, \tag{7.1d}$$

where a_c denotes the radius of a 2D circular baffle (see Fig. 7.1). However, as it will be shown in Section 7.3, these assumptions lead to strong discrepancies between the 2D vowel vocal tract transfer functions and the 3D ones. It is possible, however, to tune the 2D parameters to minimize the differences. It will be next shown how $a_{2D}(x)$, $\mu_{2D}(x)$, $Q_{2D}(t)$ and a_c, together with an additional parameter δ' can be adjusted to do so. δ' represents the length of a small duct coupled at the mouth cross-sectional area to improve 2D radiation, see Fig. 7.1.

7.2.2 Step 1: 2D vocal tract geometry

In the first step the 2D vocal tract geometry is modified to adjust the 3D vocal tract formant locations. Consider non-bent 3D and 2D vocal tracts of length L, generated from continuous area functions $S(x)$. Each vocal tract can be split into N differential ducts of constant cross-section, so that the interval $[0, L]$ becomes discretized as $0 \equiv x^1 < \cdots < x^i < x^{i+1} < \cdots < x^N \equiv L$, $\Delta x = x^{i+1} - x^i$ being the spatial step size and x^1, x^N respectively denoting the glottal and lip positions. The discretized vocal tracts can be viewed as a chain of expansion chambers with expansion ratios

$$m^i_{\mathtt{3D}} = \frac{S^i_{\mathtt{3D}}}{S^{i+1}_{\mathtt{3D}}} = \left[\frac{a^i_{\mathtt{3D}}}{a^{i+1}_{\mathtt{3D}}}\right]^2, \quad m^i_{\mathtt{2D}} = \frac{S^i_{\mathtt{2D}}}{S^{i+1}_{\mathtt{2D}}} = \frac{a^i_{\mathtt{2D}}}{a^{i+1}_{\mathtt{2D}}}, \tag{7.2}$$

for $i = 1, 2, \cdots N-1$ and with a^i_{3D}, a^i_{2D} standing for the cross-section radii. It is explained in Boij and Nilsson (2003) that in the case of expansion chambers, 2D numerical simulations compare well to 3D experimental results in the plane wave regime, whenever the 2D expansion ratio equals the 3D one. Imposing this condition for every subsequent differential duct in (7.2), the following recursive equation is found

$$a^{i+1}_{\mathtt{2D}} = a^i_{\mathtt{2D}} \left[\frac{a^{i+1}_{\mathtt{3D}}}{a^i_{\mathtt{3D}}}\right]^2 = a^i_{\mathtt{2D}}/m^i_{\mathtt{3D}}, \tag{7.3}$$

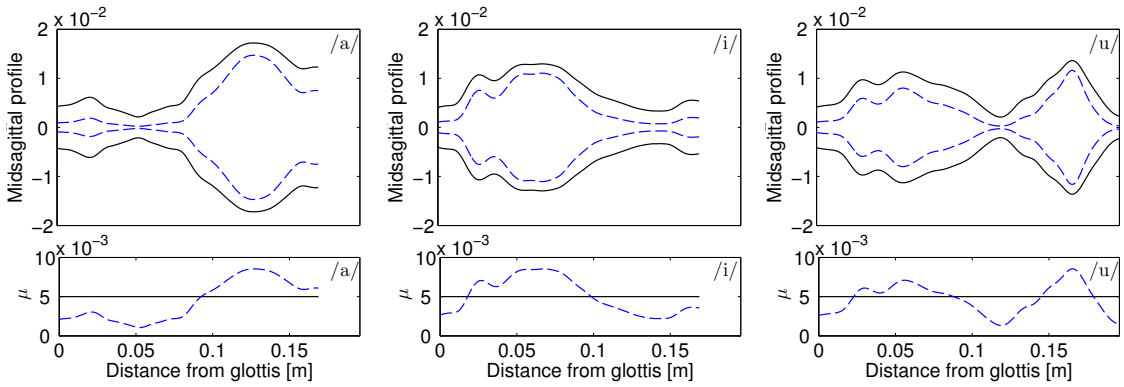
which allows building a 2D geometry having the 3D expansion ratio, once the radius $a^d_{\mathtt{2D}}$ at a given arbitrary point x^d, with $0 \leq x^d \leq L$, is fixed. For a known $a^d_{\mathtt{2D}}$, Eq. (7.3) can be rearranged as

$$a^i_{\mathtt{2D}} = \begin{cases} a^{i+1}_{\mathtt{2D}}\ m^i & 1 \leq i < d \\ a^d_{\mathtt{2D}} & i = d \\ a^{i-1}_{\mathtt{2D}}\ m^i & d < i \leq N \end{cases}, \tag{7.4a}$$

with

$$m^i = \begin{cases} S^i_{\mathtt{3D}}/S^{i+1}_{\mathtt{3D}} & i < d \\ S^i_{\mathtt{3D}}/S^{i-1}_{\mathtt{3D}} & i > d \end{cases}. \tag{7.4b}$$

It then remains to prescribe a value for the point x^d and its corresponding radius $a^d_{\mathtt{2D}}$. This can be done by controlling the onset of non-planar modes in the 2D vocal tract. The latter will be first induced at the largest cross-section of the vocal tract, so that $a^d_{\mathtt{3D}}$ and $a^d_{\mathtt{2D}}$ are chosen to be the radii precisely corresponding to these cross-sections. Consequently, x^d denotes the distance from the glottis where the maximum radius of the vocal tract is



(a) Vowel /a/ (b) Vowel /i/ (c) Vowel /u/

Fig. 7.2: Midsagittal profiles (top) and boundary admittance coefficients μ (bottom) for vowels /a/, /i/ and /u/. (— 3D, - - 2D).

achieved. Taking into account that the frequency value of the first non-planar mode in a 3D and a 2D duct can be respectively computed as (see e.g., Fletcher and Rossing, 1988)

$$f_c^{\mathsf{3D}} = \frac{1.84}{\pi}\frac{c_0}{2a^d_{\mathsf{3D}}}, \quad f_c^{\mathsf{2D}} = 0.5\frac{c_0}{2a^d_{\mathsf{2D}}}, \tag{7.5}$$

$f_c^{\mathsf{2D}} = f_c^{\mathsf{3D}}$ can be imposed providing a reasonable way to obtain the radius value a^d_{2D} to be inserted in (7.4a), from the maximum 3D one,

$$a^d_{\mathsf{2D}} = a^d_{\mathsf{3D}}\frac{0.5\pi}{1.84}. \tag{7.6}$$

In the top row of Fig. 7.2 the 3D vocal tract profiles for vowels /a/, /i/ and /u/ are given, together with their 2D counterparts obtained through the above described process. Given that $0.5\pi/1.84 < 1$ in (7.6), the 2D vocal tracts will always be narrower than the 3D versions.

7.2.3 Step 2: 2D wall losses

The second step aims at introducing the same amount of dissipation in the 2D vocal tracts walls as that encountered in the 3D ones, to adjust the formant bandwidths. To do so, the complex axial wave number k_z within the vocal tract must be the same in both cases. For a 3D cylindrical duct of radius a_{3D} with axisymmetry, the axial wave number k_z can be related to the boundary admittance coefficient μ_{3D} through

$$k_z^{\mathsf{3D}} = \sqrt{k_0^2 - k_r^2} \simeq k_0\sqrt{1 - j\frac{2\mu_{\mathsf{3D}}}{k_0 a_{\mathsf{3D}}}}, \tag{7.7}$$

where k_r is the radial wave number and $j = \sqrt{-1}$. Besides, for a 2D duct of radius $a_{\mathtt{2D}}$, the axial wave number $k_z^{\mathtt{2D}}$ is given by

$$k_z^{\mathtt{2D}} = \sqrt{k_0^2 - k_x^2} \simeq k_0 \sqrt{1 - j\frac{\mu_{\mathtt{2D}}}{k_0 a_{\mathtt{2D}}}}, \tag{7.8}$$

with k_x standing for the wave number in the x direction (perpendicular to the axial direction). The above expressions can be derived from the transcendental equations of ducts with compliant walls in Munjal (1987) (see Eq. (1.164) in pp. 36 for circular ducts and Eq. (1.124) in pp. 28 for rectangular ducts), after approximating the Bessel functions and trigonometric functions in the expressions to first order, i.e. assuming a plane wave propagation regime. Equating the axial wave numbers (7.7) and (7.8) in each differential vocal tract duct allows computing the 2D boundary admittance as

$$\mu_{\mathtt{2D}}^i = \mu_{\mathtt{3D}}^i \frac{2a_{\mathtt{2D}}^i}{a_{\mathtt{3D}}^i}. \tag{7.9}$$

If the 3D boundary admittance is set to the usual value $\mu_{\mathtt{3D}}^i = 0.005\ \forall i$ (e.g., Vampola *et al.*, 2008a) and use is made of $a_{\mathtt{3D}}^i$ and $a_{\mathtt{2D}}^i$ from the preceding Section 7.2.2 in Eq. (7.9), $\mu_{\mathtt{2D}}^i$ can be computed for the three corner vowels as shown in the bottom row of Fig. 7.2. As observed, the 2D boundary admittance is not constant in space because neither is the ratio $a_{\mathtt{2D}}^i / a_{\mathtt{3D}}^i$ in (7.9) (see also top row of Fig. 7.2).

7.2.4 Step 3: 2D glottal source

In a three-dimensional numerical simulation the particle velocity $u_{g,\mathtt{3D}}$ generated by the vocal folds can be introduced on its boundary as (see Chapter 2)

$$\nabla p \cdot \boldsymbol{n} = -\rho_0 \partial_t u_{g,\mathtt{3D}}, \tag{7.10}$$

where p is the acoustic pressure, $\boldsymbol{n}$ the normal vector pointing outwards and ∂_t the partial time derivative. Typically glottal models provide a 3D volume velocity $Q_{g,\mathtt{3D}}$, which is related to $u_{g,\mathtt{3D}}$ as

$$u_{g,\mathtt{3D}} = \frac{Q_{g,\mathtt{3D}}}{S_{g,\mathtt{3D}}}, \tag{7.11}$$

with $S_{g,\mathtt{3D}}$ being the 3D glottal cross-sectional area. In the third step, in order to recover 3D amplitude pressure levels from 2D simulations, the particle velocity $u_{g,\mathtt{2D}}$ has to be the same than in 3D. Therefore, from Eq. (7.11) it follows that

$$u_{g,\mathtt{2D}} = \frac{Q_{g,\mathtt{3D}}}{S_{g,\mathtt{3D}}} = \frac{Q_{g,\mathtt{3D}}}{\pi a_{g,\mathtt{3D}}^2}, \tag{7.12}$$

with $a_{g,\mathtt{3D}}$ denoting the radii of the 3D glottal cross-section.

7.2.5 Step 4: 2D radiation model

The goal of the fourth and final step is to make 2D radiation effects resemble the 3D ones. These effects can be quantified by means of the radiation impedance. 3D radiation effects can be complex, specially when realistic human head geometries are considered. However, it has been shown that good results for vowel production can be obtained when replacing a realistic human head by a spherical baffle of identical volume, at least from a perceptual point of view (Arnela *et al.*, 2013). This means that the analytical model for the radiation of a baffle set in a sphere can be used to a rather good extent to mimic 3D radiation effects. This theoretical model will be used as the reference for 3D, whereas as exposed in Section 7.2.1, the piston set in a circle will be used as basis for 2D. The corresponding analytical expressions can be found in Appendix A.

The idea is then to tune the 2D normalized radiation impedance $Z_r^{\mathtt{2D}}$ to the 3D one, $Z_r^{\mathtt{3D}}$, in order to get a similar radiation behavior. An optimization process to do this will be detailed in the following, which minimizes the distance between $Z_r^{\mathtt{2D}}$ and $Z_r^{\mathtt{3D}}$ by tuning the values of the circle radius a_c, and the length δ' of the short duct coupled at the end of the 2D vocal tract (mouth), in Fig. 7.1. Note that these are the only free parameters at this point given that $a_p^{\mathtt{3D}}$ and a_s are determined by the vowel to be generated, and $a_p^{\mathtt{2D}}$ has been fixed in the first step of the approach, see Section 7.2.2.

The proposed cost function to be minimized with respect to a_c and δ' is given by the relative error

$$\epsilon(a_c, \delta') = \frac{\sum_i \sqrt{(R_r^{\mathtt{2D}} - R_r^{\mathtt{3D}})^2 + (X_r^{\mathtt{2D}} - X_r^{\mathtt{3D}})^2}}{\sum_i \sqrt{(R_r^{\mathtt{3D}})^2 + (X_r^{\mathtt{3D}})^2}}, \tag{7.13}$$

where $R_r(f)$ and $X_r(f)$ respectively correspond to the radiation resistance and reactance ($Z_r = R_r + jX_r$), and the summation on index i is made over all frequency values. To first order, it is well-known that modifying the reactance results in shifting of the formant locations, whereas changing the resistance affects the formant bandwidths and pressure amplitudes. Therefore, the error function in Eq. (7.13) accounts for a balance of these effects. However, many other options could have been chosen depending on the particular goal to be reached. For instance, one could have focused on the sole optimization of X_r to better fit the formant locations, at the assumed price of getting worse results for the formant bandwidths and pressure levels.

The expression for the impedance $Z_r^{\mathtt{3D}}$ to be used in Eq. (7.13) is directly given in Appendix A. In what concerns $Z_r^{\mathtt{2D}}$, it corresponds to the input impedance of the duct

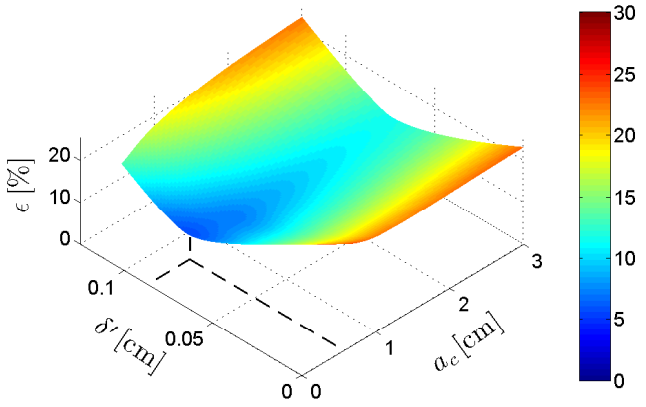


Fig. 7.3: Relative error function $\epsilon(a_c, \delta')$ in the optimization process of vowel /i/ 2D radiation impedance. a_c stands for the radius of the circle and δ' for the length of the duct coupled at the mouth exit. Dashed lines indicate the values of a_c and δ' corresponding to the minimum of ϵ.

having length δ' coupled at the mouth. This can be obtained as

$$Z_r^{2D} = \frac{1 + \mathcal{R}_p^{2D} e^{-jk\delta'}}{1 - \mathcal{R}_p^{2D} e^{-jk\delta'}}, \tag{7.14}$$

with $\mathcal{R}_p^{2D}$ standing for the reflection coefficient of a 2D piston set in a circle. This is related to the 2D piston impedance Z_p^{2D} provided in Appendix A through

$$\mathcal{R}_p^{2D} = \frac{Z_p^{2D} - 1}{Z_p^{2D} + 1}. \tag{7.15}$$

It is to be noted that use is made of the so called effective piston radius in the analytical series developments for Z_r^{3D} in Appendix A, instead of the physical one a_p^{3D}. For a sphere of radius a_s, and mouth aperture angle referenced to the center of the sphere T, the former is given by $2a_s \sin(T/2)$ whilst $a_p^{3D} = a_s \sin(T)$ (see Morse and Ingard, 1968). It is clear that both magnitudes coincide for small values of T and they have been assumed equal to simplify the optimization process. An analogous reasoning applies to Z_p^{2D} in the 2D case. Actually, it is to be noted that the theoretical developments for Z_r^{3D} and Z_p^{2D} are only valid for small to moderate values of T. This implies that a constraint on the minimum allowable value for a_c in the optimization of (7.13) has to be imposed as it could happen, for instance, that $a_c \to a_p^{2D}$, which would involve large values of T. Numerical experiments have revealed that the theoretical models perform well whenever $T < 20°$, which corresponds to demand that $a_c > 2.92 a_p^{2D}$ in the optimization process.

Table 7.1: Parameters in Z_r^{3D} and Z_r^{2D} and results of the optimization process for the 2D radiation impedance. a_p is the piston radius (mouth aperture), a_s the sphere radius, a_c the circle radius, δ' the length of the coupled duct, and T is the mouth aperture angle referenced to the sphere or circle center. All values in cm.

	Z_r^{3D}			Z_r^{2D}			
	a_p^{3D}	a_s	T	a_p^{2D}	a_c	δ'	T
/a/	1.226	9	7.83°	0.745	2.18	0.063	19.99°
/i/	0.538	9	3.43°	0.191	0.57	0.087	19.59°
/u/	0.226	9	1.44°	0.032	1.16	0.067	1.58°

As an example, the cost function $\epsilon(a_c, \delta')$ has been plotted for vowel /i/ in Fig. 7.3. As observed from the comparison between the section curve with $\delta' = 0$ and that passing through the minimum, the inclusion of the additional short duct greatly improves the results. The optimum pairs $[a_c, \delta']$ obtained for each vowel have been summarized in Table 7.1. Note in all cases that the angle T is always less than 20° (actually the constraint $T < 20°$ is only necessary for vowel /a/ in the optimization process). Also observe in Table 7.1 that the obtained values for a_c are much smaller than a_s. An intersection between the vocal tract geometry and the circular baffle is then expected since $2a_c < L$ (see Fig. 7.4), but as it will be explained in the following subsection, this will not represent a problem for the FEM simulations.

7.2.6 Time domain finite element simulations

The finite element method (FEM) has been used to solve the acoustic wave equation within the 2D and 3D computational domains Ω. These are made of the vocal tract plus an external space surrounding the human head (approximated by spherical or circular baffles) with dimensions $0.28 \times 0.24 \times 0.28$ m in 3D and 0.28×0.24 m in 2D. In both cases the external space is surrounded by a perfectly matched layer (PML) of width $L = 0.1$ m to absorb outgoing sound waves, and configured to get a reflection coefficient of $r_\infty = 10^{-4}$. Details on the used FEM formulation with PML can be found in Arnela and Guasch (2013).

The computational domains Ω have been meshed using unstructured tetrahedral meshes for 3D and triangular meshes for 2D. The mesh size h for the 3D/2D simulations ranges

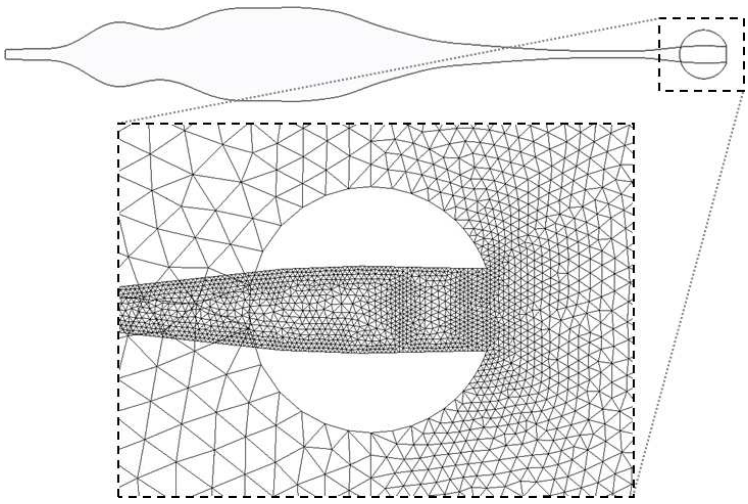
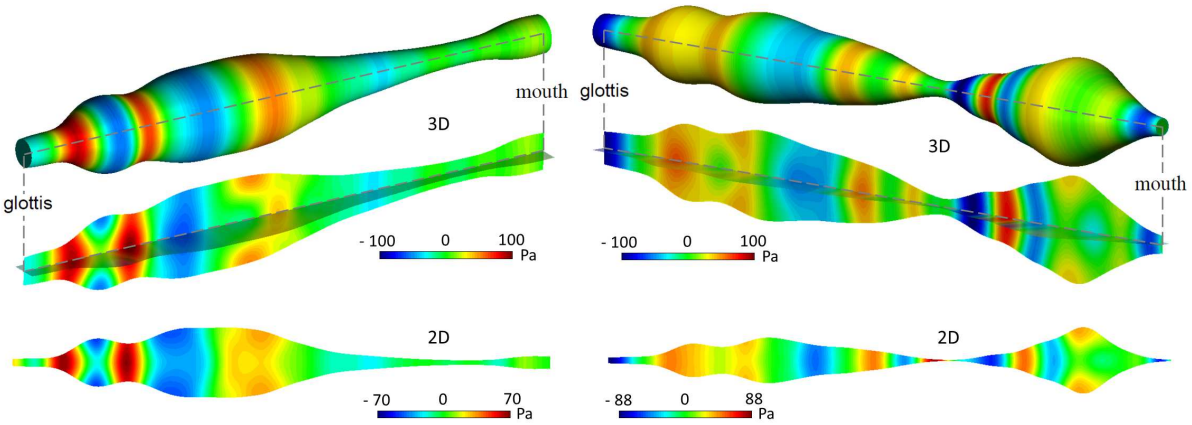


Fig. 7.4: Detail of the 2D mesh for vowel /i/ around the mouth aperture.

from $h = 0.002/0.0002$ m within the vocal tract and the immediate area outside the mouth, $h = 0.005/0.002$ m in the outer space, and $h = 0.0075/0.002$ m in the PML. The smaller mesh sizes within the 2D the vocal tract are necessary because of the strong constrictions in the modified 2D geometries. Besides and due to the high computational cost of 3D simulations, the mesh size in free-field space has been increased to $h = 0.005$ m in the outer volume and to $h = 0.075$ m in the PML. This limits the frequency analysis of the 3D results to ~ 10 kHz (following the 10 nodes per wavelength criterion, see e.g., Ihlenburg, 1998), while it reaches 20 kHz for the 2D case.

The above mesh configurations result, approximatively, in 7 million and 0.15 million elements for the 3D and 2D computational domains, respectively. Note that although smaller element sizes are used for 2D, the total number of elements is strongly reduced from 3D to 2D. In Fig. 7.4 some 2D mesh details for vowel /i/ are presented. As a consequence of the optimization process (see Section 7.2.5) the circular baffle has a smaller radius than the vocal tract length. However, this poses no problem in terms of the computation, since the nodes outside the vocal tract are disconnected from those inside. Therefore, sound waves leaving the vocal tract will be scattered by the baffle and will propagate outwards without difficulty.

With regard to the boundary conditions, a Gaussian pulse (Takemoto *et al.*, 2010) has been used for the volume velocity $Q(t)$ and imposed at the glottal cross-sectional area. The pulse has been low pass filtered to avoid the appearance of high frequency numerical errors. A constant boundary admittance coefficient $\mu_{3D} = 0.005$ has been imposed at the 3D vocal tract walls, whereas for the 2D simulations μ_{2D} is space varying and has been obtained from Eq. (7.9). For both, the spherical and circular baffles, hard walls ($\mu = 0$)



(a) Vowel /i/ (b) Vowel /u/

Fig. 7.5: Snapshot of the acoustic pressure for vowels /i/ and /u/ at time instant $t = 1.15$ ms and $t = 1.675$ ms, respectively. (top) 3D vocal tract (wall surface, midsagittal and midtransverse planes). (bottom) 2D-proposed vocal tract.

have been assumed. For the time evolution, a speed of sound of c_0=350 m/s and an air density of $\rho_0 = 1.14$ kg/m^3 have been prescribed. The sampling rates chosen for the 3D and 2D simulations are $f_s^{3D} = 1/\Delta t = 2000$ kHz and $f_s^{2D} = 4000$ kHz, which ensure that the corresponding stability conditions of the Courant-Friedrich-Levy type required by explicit numerical schemes are fulfilled. Note that a higher sampling rate f_s is needed in 2D because, as said above, very small elements are necessary to mesh the constrictions of the modified 2D geometries. 50 ms events have been simulated in a serial computing system with processor Intel(R) Core(TM) i5 2.8 GHz (see Table 7.4 for computational times). In Fig. 7.5, a snapshot of the vocal tract acoustic pressure of vowel /i/ at the time instant $t = 1.15$ ms is presented, for a 3D reference case and for the proposed 2D approach.

Finally, the vocal tract transfer functions for the 3D and 2D numerical simulations have been computed as

$$H(f) = \frac{P_o(f)}{Q(f)}, \tag{7.16}$$

with $P_o(f)$ and $Q(f)$ corresponding to the Fourier Transforms of the acoustic pressure $p_o(t)$ at the mouth and the volume velocity $Q(t)$ input at the glottis. For a fair comparison with 1D models in subsequent sections, $p_o(t)$ is captured inside the vocal tract, at a distance of 0.003 m from its exit (this distance does not include the coupled duct with length δ', which is treated as a radiation element).

7.3 Results

Results for vowels /a/, /i/ and /u/ using the proposed methodology are next presented. The 2D simulations carried out following the above described approach will be compared to a 3D reference case, to the standard 2D method that considers a midsagittal cut of the 3D vocal tract, and to a 1D frequency domain technique based on standard chain matrices for transmission lines (similar to that in Sondhi and Schroeter, 1987, but using the same wall losses as in the 3D reference case and the radiation impedance of a piston set in a sphere). For brevity, the above approaches will be hereafter designated as the 3D-case, the 2D-proposed case, the 2D-standard case and the 1D-case. It should be noted that the comparisons with the 1D model are somewhat unfair given that this is a frequency domain model, whereas all other model results come from time domain simulations. Imposing free field radiation in time domain 1D models is a rather intricate task and the inclusion of wall damping has to be done through the inclusion of an extra damping term in the equation, which combines spatial and time partial derivatives (see e.g., Doel and Ascher, 2008). As it is possible to implement the 3D wall and radiation losses in the 1D frequency model, this has at least the advantage that the observed differences with other models could be basically attributed to non-planar wave propagation.

In order to check each step of the proposed methodology (step 1: vocal tract profile, step 2: wall losses, step 3: glottal flow and step 4: radiation), two different situations have been considered. In the first one no radiation losses are considered and the computational domain is ended at the exit of the vocal tract. An homogenous Dirichlet boundary condition $p = 0$ is imposed on it to emulate an open-end termination. This open-end occurrence is intended to check the performance of steps 1, 2 and 3. In the second situation, radiation takes place as described in Section 7.2.6 and the efficiency of the 2D radiation model in step 4 is evaluated individually. The computational times for both situations are summarized in Table 7.4 for the 3D-case, the 2D-proposed and the 1D-case. The model accuracies will be contrasted to computational efficiency in what follows.

7.3.1 Open-end conditions

Figure 7.6 shows the obtained vocal tract transfer functions (VTTF) for the open-end situation up to $f_{max} = 20$ kHz, Table 7.2 the formant locations and deviations in % with respect to the 3D-case, and Table 7.3 the formant bandwidths. Step 3 (compensation of glottal flow) has been always applied in the 2D models (including the 2D-standard) for an easier comparison between cases, since it simply produces an offset variation of the VTTFs. Step 2 has been first disabled to check the efficacy of the proposed vocal tract

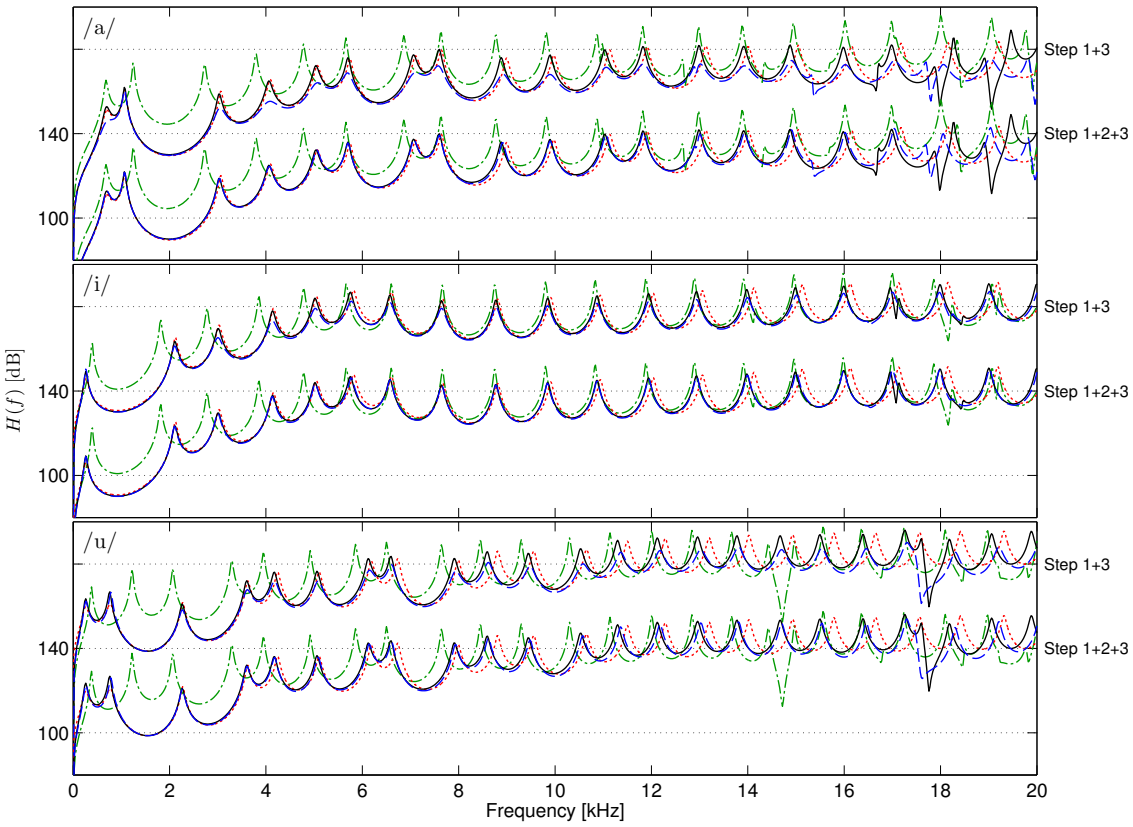


Fig. 7.6: Vocal tract transfer functions $H(f)$ for vowels /a/, /i/ and /u/ when $p = 0$ is imposed at the mouth cross-sectional area (open-end conditions). Step 1 (vocal tract profile), step 2 (boundary admittance) and step 3 (glottal flow) have been used to construct the 2D-proposed model. (— 3D-case, - - 2D-proposed, —· 2D-standard, · · · 1D-case).

profiles in step 1, and then turned on to observe if proper wall losses are recovered.

First the 2D-standard case will be compared to the 3D-case in Fig 7.6. Strong discrepancies in the formant locations are observed, specially for the first formants at low frequencies. This will result in large audible differences when synthesizing a vowel, since the deviations with respect to the 3D-case (see Table 7.2) are clearly larger than 3-5 % for some of the formants (Flanagan, 2008). Moreover, the peaks are sharper in 2D than in 3D (see also Table 7.3). Having a look at Eq. (7.9), this may be attributed to the fact that to get the same amount of wall absorption for 2D than for the 3D-case, a factor 2 should have been applied to the boundary admittance coefficient, given that if the midsagittal 3D cut is taken for the 2D vocal tract the ratio $a_{2D}/a_{3D} = 1$.

In contrast to the 2D-standard results, when step 1 is used to construct the vocal tract profiles very close formant locations to the ones in the 3D-case are obtained up to 10 kHz

(see dashed line in Fig. 7.6 and Table 7.2). Beyond 10 kHz, higher order modes appear in the VTTFs which apparently cannot be correctly reproduced by any of the two 2D approaches. Yet some of them are recovered in the 2D-proposed case, like the deep around 18 kHz for vowels /a/ and /u/. The inclusion of step 2 (wall losses) allows to get proper bandwidths (see Table 7.3), otherwise some large disparities can be appreciated, see e.g., the first formant of vowel /a/ in Fig. 7.6. Actually, it is apparent that one could not expect to regain all 3D higher order acoustic modes from 2D simulations, only some of them can be obtained. However, and as mentioned in the introduction, 2D simulations seem to produce more natural sounds than those arising from 1D simulations, even if high order 3D modes are not perfectly matched, which make them appealing e.g., for articulatory speech synthesis (see e.g., Howard *et al.*, 2009; Mullen *et al.*, 2006).

When testing the 1D-case against the 2D-proposed and the 3D-case, results only compare well for low frequencies, as expected, as no high order modes can be generated in 1D (departures can be clearly observed beyond 5-6 kHz for vowels /i/ and /u/). Besides, some formant deviations to higher frequencies can be also appreciated for moderate frequency values, see e.g. the fifth formant of vowel /u/. This is so because in 2D and 3D front waves can curve at large area discontinuities (see Fig. 7.5). This can be understood as an artificial lengthening of the differential duct where the discontinuity takes place (see e.g., Kang and Ji, 2008), inducing the shift to the left of the 2D and 3D formants; a phenomenon that obviously cannot naturally occur in the 1D case. Note that the initial condition in Eq. (7.6) for the vocal width guarantees that this effect can take place. 1D models can emulate this lengthening phenomena by introducing inner-length corrections at the sudden expansions/constrictions of the vocal tract (Sondhi, 1983). Though for the herein analyzed vowels this effect does not seem to be influential (the differences between 1D and 2D/3D Table 7.2 are less than 3-5%), the phenomenon may be of importance in other situations and actually, some current 1D articulatory speech synthesizers take it into account to improve the formant accuracy (see e.g. Birkholz, 2013). In what concerns formant bandwidths, a general matching is observed between 1D and 3D (also 2D), although some formants in the low frequency region are found overdamped in 1D (see e.g. F_1 and F_2 for vowel /u/ in Table 7.3), which might be perceptually relevant.

With regard to the computational times in Table 7.4, it can be observed that those of 2D simulations are much lower than those of 3D simulations. Though in the present work it has not been attempted to reach real-time computation (the simulated time interval was 50 ms for 2D and 3D, see Section 7.2.6), it has been shown that real-time applications can be reached for both 1D (see e.g., Doel and Ascher, 2008) and 2D simulations (see e.g, Mullen *et al.*, 2007), using similar configurations to those presented in this section.

Table 7.2: First formant locations F_i (Hz) and formant deviations $\Delta F_i/F_i$ (%) with respect to the 3D-case in the 2D-proposed (2Dp), the 2D-standard (2Ds) and the 1D-case for vowels /a/, /i/ and u/. Open-end conditions ($p = 0$) are considered at the mouth cross-section.

Vowel	Case	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
/a/	3D	696	1068	3031	4064	5039	5703	7065	7595
	2Dp	701	1069	3036	4071	5057	5705	7072	7598
	2Ds	682	1249	2730	3793	4780	5661	6862	7632
	1D	705	1072	3065	4091	5066	5744	7108	7642
/i/	3D	263	2111	3010	4138	5019	5753	6575	7645
	2Dp	266	2106	3003	4131	5024	5772	6578	7647
	2Ds	392	1816	2779	3855	4784	5670	6605	7664
	1D	262	2121	3029	4158	5048	5820	6606	7668
/u/	3D	259	757	2264	3603	4173	5062	6126	6594
	2Dp	264	783	2266	3609	4179	5085	6144	6610
	2Ds	378	1224	2068	3292	3950	4921	5862	6504
	1D	260	769	2275	3624	4244	5105	6186	6736

Vowel	Case	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7	Δ_8
/a/	2Dp	0.72	0.09	0.16	0.17	0.36	0.04	0.1	0.04
	2Ds	-2.01	16.95	-9.93	-6.67	-5.14	-0.74	-2.87	0.49
	1D	1.29	0.37	1.12	0.66	0.54	0.72	0.61	0.62
/i/	2Dp	1.14	-0.24	-0.23	-0.17	0.1	0.33	0.05	0.03
	2Ds	49.05	-13.97	-7.67	-6.84	-4.68	-1.44	0.46	0.25
	1D	-0.38	0.47	0.63	0.48	0.58	1.16	0.47	0.3
/u/	2Dp	1.93	3.43	0.09	0.17	0.14	0.45	0.29	0.24
	2Ds	45.95	61.69	-8.66	-8.63	-5.34	-2.79	-4.31	-1.36
	1D	0.39	1.59	0.49	0.58	1.7	0.85	0.98	2.15

Table 7.3: First formant bandwidths BW_i (Hz) for the 3D-case, the 2D-proposed (2Dp), the 2D-standard (2Ds) and the 1D-case for vowels /a/, /i/ and u/. Open-end conditions ($p = 0$) are considered at the mouth cross-section.

Vowel	Case	BW_1	BW_2	BW_3	BW_4	BW_5	BW_6	BW_7	BW_8
/a/	3D	135	50	87	117	117	83	115	89
	2Dp	137	51	86	120	115	87	113	93
	2Ds	58	34	49	55	54	45	54	48
	1D	126	59	76	107	111	84	114	86
/i/	3D	57	82	101	97	105	97	94	93
	2Dp	58	82	101	97	105	101	94	92
	2Ds	33	46	52	50	50	47	47	47
	1D	95	71	92	90	100	97	93	92
/u/	3D	69	62	84	94	93	95	99	90
	2Dp	70	64	84	95	93	98	100	89
	2Ds	38	38	45	51	47	47	51	46
	1D	125	122	76	84	87	93	90	86

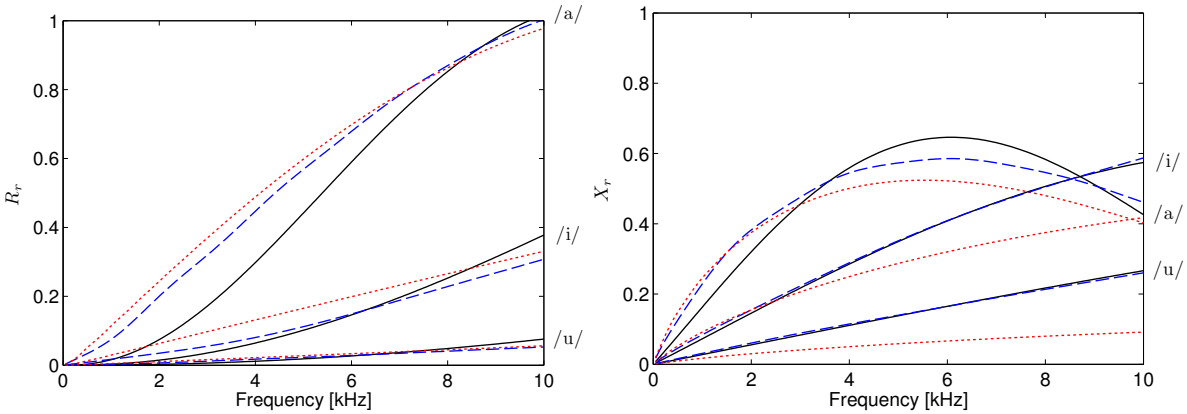
Table 7.4: Computational times for the 1D-case, the 2D-proposed and the 3D-case when open-end boundary conditions are used ($p = 0$) on the mouth section and when radiation is allowed. Remark: 1D is not directly comparable to 2D and 3D simulations because it is a frequency domain model while 2D and 3D work in the time domain.

	Open-end			Radiation		
	1D	2D	3D	1D	2D	3D
/a/	5 s	4 min	44 min	8 s	2.7 h	56 h
/i/	5 s	4 min	30 min	10 s	2.7 h	59 h
/u/	5 s	4 min	31 min	13 s	2.7 h	57 h

7.3.2 Radiation conditions

In order to check the performance of the radiation optimization in the fourth step of the method, two different models have been considered. The first one is made of a 2D vocal tract that has been tuned according to steps 1, 2 and 3. The vocal tract is embedded in a circle consisting of the midsagittal cut of the 3D-case sphere. In other words, the values $a_c = a_s$ and $\delta' = 0$ are taken for it. The second model consists of the same 2D vocal tract but with step 4 also being implemented. That is to say the values for a_c and δ' in Table 7.1 have been used in this occasion. For brevity, the two models will be respectively refereed to as the 2D-proposed case with and without optimization. Results will be presented for them and also compared with those of the 3D-case with free-field radiation conditions and those of the 1D-case with the piston-set on a sphere load radiation model implemented on it. The 2D-standard case will be left aside due to its poor performance when compared to others. The analysis will be restricted up to 10 kHz due to the high computational cost of the free field 3D simulations.

Let us first comment on the 2D theoretical radiation impedances obtained that can be computed using the values arising from the optimization process (see Section 7.2.5), and compare them to the 3D ones. In Fig. 7.7, the 3D and 2D (with and without optimization) impedance curves have been plotted in terms of the radiation resistance R_r (Fig. 7.7a) and the reactance X_r (Fig. 7.7b). With regard to the former, both the 2D case with and without optimization present similar values to those of the 3D-case for vowel /u/. In



(a) (b)

Fig. 7.7: Radiation impedance Z_r for vowels /a/, /i/ and /u/ computed from theoretical models. (a) Radiation resistance R_r and (b) Radiation reactance X_r, $Z_r = R_r + jX_r$. (— 3D-case, - - 2D-proposed with optimization, ··· 2D-proposed without optimization).

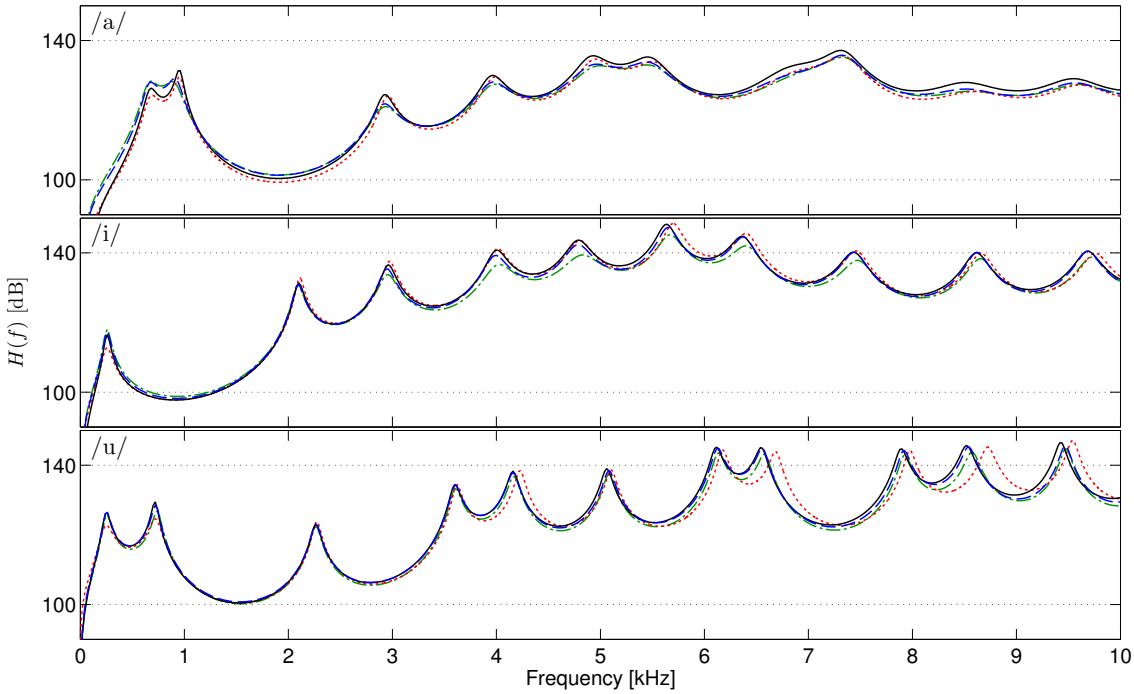


Fig. 7.8: Vocal tract transfer functions $H(f)$ for vowel /a/, /i/ and /u/ considering radiation. (— 3D-case, - - 2D-proposed with optimization in the radiation model, –· 2D-proposed without optimization, · · · 1D-case).

contrast, for vowel /i/ differences become apparent and the 2D-case with optimization better approximates the 3D curve for most of the spectrum; see Fig. 7.7a. For vowel /a/, both 2D models clearly depart from the 3D resistance, though slightly better results are obtained for the case with optimization. In what concerns the reactance, a striking agreement is found between the 3D-case and the 2D with optimization for vowels /u/ and /i/ (see Fig. 7.7b), whereas this is not the case for the 2D-case without optimization. The results are not as good for vowel /a/, yet the 2D-case with optimization performs better than that without optimization.

The impedance plots shown in Fig. 7.7 are useful to understand the differences between the 3D and 2D VTTFs shown in Fig. 7.8. As mentioned in previous sections, increasing the reactance is well known to shift the formant locations downward in frequency. This is what is actually observed for vowels /i/ and /u/ in Fig. 7.8, where the 2D-case with optimization better fits the 3D-case VTTF. The 2D with no optimization has lower reactance (see Fig. 7.7b) and therefore its formants are shifted to the right. Note that some exceptions occur, like the third formant of vowel /u/ which is a back-cavity formant (see e.g. Apostol *et al.*, 2004) and therefore becomes less influenced by changes in radiation (see e.g., Stevens, 2000). Besides, the discrepancies already found between $9 - 10$ kHz

for vowel /u/ in Fig. 7.6, are also observed in Fig. 7.8. No significative deviations are observed in the VTTF of vowel /a/ despite the 2D reactance values did not fit so well the 3D ones.

Increasing resistance generally results in larger formant bandwidths and lower amplitudes (with dependence on the formant cavity affiliation phenomenon). A look at vowel /u/ in Fig. 7.8 reveals that very good results are obtained for both 2D models, given that they closely match the 3D resistance (see Fig. 7.7a). Only a small decrease of the 2D pressure levels is observed beyond $\sim$ 5 kHz, because the 2D resistance also diminishes slightly with respect to the 3D one at these frequencies. In contrast, for vowel /i/ better results are obtained for the case with optimization (see Fig. 7.8 beyond $\sim$ 3 kHz), the formants being smoothed if this is not carried out. For example, the fourth and fifth formants that lie between 4 and 5 kHz have larger bandwidths since their resistance is also higher at this frequency range (see Fig. 7.7a). The discrepancies between 3D and 2D VTTFs become more visible for vowel /a/ due to the lower 3D resistance; see Fig. 7.7a. For example, the first 3D-case formant of vowel /a/ not only has a smoother shape but also less amplitude level. Close inspection reveals, however, some slight improvements in some formants (e.g., second and third) for the 2D case with optimization with respect to the 2D case without optimization.

Results for the 1D-case have been also presented in Fig. 7.8. As expected, values similar to the 3D-case ones are obtained, since the equivalent theoretical radiation impedance (piston set on a sphere) has been applied to emulate radiation in the 1D model. Differences in formant locations can be basically appreciated for vowel /u/ above 4 kHz, the explanation for this behavior already being given in Section 7.3.1 and not related to radiation.

In what concerns the computational times, the inclusion of the free-field space in 2D and 3D obviously has noticeably increased the total duration of the simulations (see Table 7.4). The computational times for the 2D case are still much smaller than the 3D ones and they could actually be reduced as the comparison is somewhat unfair in this case. As commented in Section 7.2.6, the meshes in 2D are fine enough to reach 20 kHz while only 10 kHz can be attained for 3D. The computational time is still low for the 1D case given that it makes use of the theoretical impedance radiation model of a piston set in a sphere, detailed in Appendix A.

7.4 Conclusions

Numerical simulations show that the vocal tract transfer function (VTTF) of a 2D vocal tract, generated from the midsagittal cut of a 3D one, presents strong discrepancies in terms of formant locations, bandwidths and amplitudes when compared to a 3D VTTF. Alternatively, a four step methodology has been proposed for tuning 2D models so as to exhibit the behavior of 3D vocal tract shapes with circular cross-sections. The method has shown substantial improvement when compared to the standard 2D approach and allows taking benefit of the great reduction in computational cost of 2D simulations, without paying strong quality sacrifices for it. The observed differences between the time-domain 3D and 2D proposed approach, and the 1D frequency-domain method are not huge, which is partially attributed to the fact that simple 3D vocal tract with circular-cross sections have been used. Moreover, larger discrepancies could have been expected if comparisons were made with respect to a 1D time-domain model, given their difficulty to mimic some phenomena such as radiation or wall losses.

2D methods can present some advantages with respect to 1D models, as non planar waves can be generated and both, radiation and wall losses, can be naturally accounted for. Moreover, one could consider using asymmetrical shapes and bent vocal tracts, which will produce the onset of new transverse modes and probably improve the naturalness of the generated sound. Complex side branches could be also included in 2D models to take into account the influence e.g., of the piriform fossae or of the nasal tract. As a result of this work, it may be concluded that accurate 2D simulations have the potential to provide a reasonable balance in terms of quality and computational cost between 3D and 1D approaches, though there is much more to be investigated concerning their future possibilities.

Part III

FEM generation of diphthongs

Chapter 8

Finite Element formulation for diphthong production

For the numerical simulation of vowel production, the irreducible wave equation for the acoustic pressure is typically solved. However, diphthong sounds require moving vocal tract geometries so that the wave equation has to be expressed in an Arbitrary Lagrangian-Eulerian (ALE) framework. It then becomes more convenient to directly work with the wave equation in its mixed form, which not only involves the acoustic pressure but also the acoustic particle velocity. In turn, this entails some numerical difficulties that require resorting to stabilized FEM approaches. In this chapter, the mixed wave equation expressed in an ALE frame of reference will be numerically solved using FEM and stabilized following a subgrid scale strategy, which will allow one to use the same interpolation for the acoustic pressure and acoustic particle velocity. Some numerical examples will be finally presented showing a good performance of the proposed approach. Diphthong sounds /ai/ and /au/ will be generated using three-dimensional vocal tracts with circular cross-sections and their equivalent tuned two-dimensional counterparts.

This chapter is based on the following works

- Guasch, O., Arnela, M., Codina, R., and Espinoza, H., "A stabilized finite element method for the mixed wave equation in an ALE framework with application to diphthong production," J. Comput. Phys., Submitted.
- Arnela, M., Guasch, O., Codina, R., and Espinoza, H. (**2014**), "Finite element computation of diphthong sounds using tuned two-dimensional vocal tracts", in *Proc. of 7th Forum Acousticum*, Kraków, Poland.

- Guasch, O., Codina, R., Arnela, M., and Espinoza, H. (**2014**), "A stabilized Arbitrary Lagrangian Eulerian Finite Element Method for the mixed wave equation with application to diphthong production", in *11th World Congress on Computational Mechanics (WCCM)*, Barcelona, Spain.

8.1 Introduction

The acoustic pressure is the magnitude of interest for numerical voice simulation. Therefore, for vowel production, the standard irreducible wave equation, or its Fourier transform, the Helmholtz equation, are the usual equations being solved (e.g., Matsuzaki and Motoki, 2007; Vampola *et al.*, 2008a; Arnela and Guasch, 2013; Arnela *et al.*, 2013). However, approaches dealing with the wave equation in mixed form that involves both, the acoustic pressure and the acoustic velocity, can also be considered (see e.g., Takemoto *et al.*, 2010; Codina, 2008). Actually, it will be shown herein that to generate diphthong sounds it precisely becomes more convenient to deal with the later. In the case of diphthongs, one has to consider moving domains. For example, if one intends to generate the diphthong /ai/, the vocal tract geometry has to smoothly change from the articulatory position of vowel /a/ to the position of vowel /i/ (see Fig. 8.1). Therefore, dynamic vocal tracts will be required and consequently it becomes necessary to express the involved equations in an Arbitrary Lagrangian-Eulerian (ALE) framework (Hughes *et al.*, 1981; Huerta and Liu, 1988; Donea *et al.*, 2004), rather than in a pure Eulerian one. This can be done more naturally for the wave equation in mixed form.

In contrast to the irreducible wave equation presented in Chapter 2, the numerical resolution of the mixed wave equation presents some additional difficulties. To solve it by FEM one has to express this equation in its weak or variational form and then discretized it. A compatibility inf-sup condition has to be satisfied for the problem to be well-posed, which in the case of the mixed wave equation is not directly inherited by the standard Galerkin FEM approach. It becomes then necessary to use different interpolations for the acoustic pressure and particle velocity, which makes it more difficult to code. Alternatively, one can resort to stabilization which allows one to use equal interpolation for both magnitudes. Different options exist in the literature (see e.g., Codina, 1998), although in this work we will adopt a residual based multiscale method (Hughes, 1995; Hughes *et al.*, 1998; Codina, 2002) for the mixed wave equation (Codina, 2008). A detailed numerical analysis reveals that in this case the solution becomes bounded by the problem data in a new energy norm that contains all designed stabilization terms (Badia *et al.*, 2014). In the case of its ALE counterpart the

situation becomes more intricate. This equation has the same mathematical structure of the linearized modified Boussinesq equation for shallow waters with convection. Numerical experiments show that the Galerkin FEM solution for the latter exhibits strong high-frequency oscillations if equal interpolations are used for the gravity waves depth and velocity (Langtangen and Pedersen, 1998; Walkley and Berzins, 2002). Again these oscillations can be overcome if one draws on stabilization strategies (Hauke, 1998; Codina *et al.*, 2008). In this work we will present a stabilized FEM approach for the ALE wave equation in mixed form based on subgrid scales, following the main lines in Codina *et al.* (2008). It is to be noted however, that contrary to the variational mixed wave equation, little is know with regard to the well-posedness of the variational continuous and discrete ALE mixed wave equation. Actually, and as far as we know, to date only very partial results have been proved showing that the continuous linearized two-dimensional shallow water equations (and by extension the ALE mixed wave equation) are only well posed for some specific boundary conditions in a very simple rectangular geometry (Huang and Temam, 2014).

This chapter is organized as follows. In Section 8.2 the wave equation is expressed in an ALE frame of reference and supplemented with appropriate boundary and initial conditions for diphthong sounds. This problem is then numerically solved with FEM in Section 8.3. First, it is expressed in its variational form (Section 8.3.1). Then, it is discretized in space following a subgrid scale strategy to use equal interpolation for the acoustic pressure and acoustic particle velocity (Section 8.3.2). Next the so called matrix of stabilization parameters need for the stabilization is derived (Section 8.3.3). The final numerical scheme including both space and time discretization is then presented (Section 8.3.4). The general methodology followed to perform a numerical simulation of a diphthong sound is described in Section 8.4. In Section 8.5 a numerical example used to check the proposed FEM approach is first shown. This consists of a two-dimensional tube of constant cross-section with time variable length due to periodic movement of the exit boundary, which can be interpreted as a very rough approximation of a vocal tract in a rest position with lip protrusion. Next, the diphthongs /ai/ and /au/ are generated as a second example using three-dimensional simplified geometries of the vocal tract with circular cross-sections. Lastly, the same diphthongs sounds are synthesized but this time using a two-dimensional approach that provides a very good balance between computational cost and voice quality. The tuned two-dimensional vocal tract model developed in Chapter 7 is extended to account for dynamic geometries. Conclusions finally close the chapter in Section 8.6.

8.2 Problem statement

8.2.1 The acoustic wave equation in an ALE framework

The linearized continuity and momentum equations for sound propagation in a stationary inviscid fluid read, in a spatial (Eulerian) frame of reference,

$$\frac{1}{\rho_0 c_0^2}\partial_t p + \nabla \cdot \boldsymbol{u} = f_p, \tag{8.1a}$$

$$\rho_0 \partial_t \boldsymbol{u} + \nabla p = \boldsymbol{f_u}, \tag{8.1b}$$

where $p(\boldsymbol{x}, t)$ and $\boldsymbol{u}(\boldsymbol{x}, t)$ respectively denote the acoustic pressure and the acoustic particle velocity and c_0 and ρ_0 stand for the speed of sound and for the air density. $f_p(\boldsymbol{x}, t)$ represents a volume source distribution and $\boldsymbol{f_u}(\boldsymbol{x}, t)$ an external body force per unit volume. The partial time derivative is denoted with ∂_t. Equation (8.1) is sometimes referred to as the wave equation in mixed form. Its standard irreducible counterpart can be easily derived after combing the mixed wave equation (8.1) as

$$\rho_0 \partial_t (8.1\text{a}) - \nabla \cdot (8.1\text{b}), \tag{8.2}$$

which yields

$$\frac{1}{c_0^2}\partial_{tt}^2 p - \nabla^2 p = \rho_0 \partial_t f_p - \nabla \cdot \boldsymbol{f_u}. \tag{8.3}$$

Solving Eq. (8.3) supplemented with appropriate boundary and initial conditions suffices for numerical vowel production (see e.g, Arnela and Guasch, 2013, and also Chapter 2), because, as said in the Introduction, the generated acoustic pressure is the sole quantity of interest. However, when facing the production of diphthongs one has to deal with moving domains. In such a situation, it becomes convenient to express the time derivative of any fluid property, say ξ, in a referential frame moving with the mesh, while keeping its spatial derivatives in an Eulerian frame (quasi-Eulerian ALE formulation, Hughes *et al.*, 1981; Huerta and Liu, 1988). To get the ALE expressions for the continuity and momentum equations in Eq. (8.1), one has to replace

$$\partial_t \xi \leftarrow \partial_t \xi - \boldsymbol{u}_{\text{dom}} \cdot \nabla \xi \tag{8.4}$$

in them, with $\boldsymbol{u}_{\text{dom}}(\boldsymbol{x}, t)$ denoting the mesh velocity. This results in the ALE wave equation in mixed form

$$\frac{1}{\rho_0 c_0^2}\partial_t p - \frac{1}{\rho_0 c_0^2}\boldsymbol{u}_{\text{dom}} \cdot \nabla p + \nabla \cdot \boldsymbol{u} = f_p, \tag{8.5a}$$

$$\rho_0 \partial_t \boldsymbol{u} - \rho_0 \boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u} + \nabla p = \boldsymbol{f_u}. \tag{8.5b}$$

Note that combining Eq. (8.5a) and Eq. (8.5b) to get an irreducible equation for the acoustic pressure is not straightforward in this case. An analogous procedure to that followed to get Eq. (8.3) should be followed. Replacing the time derivatives in Eq. (8.2) by $\partial_t \xi \leftarrow \partial_t \xi - \boldsymbol{u}_{\text{dom}} \cdot \nabla \xi$ and using instead the ALE mixed wave equation (8.5), the irreducible form of the wave equation in an ALE framework can be derived after applying

$$\rho_0 \partial_t (8.5\text{a}) - \rho_0 \boldsymbol{u}_{\text{dom}} \nabla \cdot (8.5\text{a}) - \nabla \cdot (8.5\text{b}), \tag{8.6}$$

which yields

$$\begin{aligned} \frac{1}{c_0^2} \partial_{tt}^2 p - \nabla^2 p - \frac{1}{c_0^2} \partial_t \left(\boldsymbol{u}_{\text{dom}} \cdot \nabla p \right) + \frac{1}{c_0^2} \boldsymbol{u}_{\text{dom}} \cdot \nabla \left(\boldsymbol{u}_{\text{dom}} \cdot \nabla p \right) - \rho_0 \boldsymbol{u}_{\text{dom}} \cdot \nabla (\nabla \cdot \boldsymbol{u}) \\ + \rho_0 \nabla \cdot (\boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}) = \rho_0 \partial_t f_p - \rho_0 \boldsymbol{u}_{\text{dom}} \cdot \nabla f_p - \nabla \cdot \boldsymbol{f}_{\boldsymbol{u}}. \end{aligned} \tag{8.7}$$

Note in Eq. (8.7) that some of the new terms that appear when expressing Eq. (8.3) in an ALE framework include the particle velocity $\boldsymbol{u}$, e.g. $\rho_0 \nabla \cdot (\boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u})$. Since this magnitude will have to be computed anyway, it seems more appropriate to solve the mixed wave equation which directly provides the pressure and velocity fields.

8.2.2 Boundary and initial conditions

The boundary and initial conditions that are used for the ALE wave equation in mixed form, Eq. (8.5), applied to our particular problem are as follows. Consider a computational domain Ω standing for the vocal tract airway. Its boundary $\partial\Omega$ can be split in three non-intersecting regions (see Fig. 8.1): Γ_{G} for the glottal cross-sectional area, Γ_{W} for the vocal tract walls, and Γ_{M} for the mouth aperture. We impose

$$\boldsymbol{u} \cdot \boldsymbol{n} = g(t) \qquad \text{on } \Gamma_{\text{G}},\ t > 0, \tag{8.8a}$$

$$\boldsymbol{u} \cdot \boldsymbol{n} = p / Z_w \qquad \text{on } \Gamma_{\text{W}},\ t > 0, \tag{8.8b}$$

$$p = 0 \qquad \text{on } \Gamma_{\text{M}},\ t > 0, \tag{8.8c}$$

$$p = 0,\ \boldsymbol{u} = 0 \qquad \text{in } \Omega,\ t = 0, \tag{8.8d}$$

where $g(t)$ is a function that describes the time evolution of the acoustic velocity generated by the vocal folds, Z_w is the impedance of the vocal tract walls, related to the boundary admittance coefficient by means of $\mu = \rho_0 c_0 / Z_w$, and $\boldsymbol{n}$ is the normal vector pointing outwards $\partial\Omega$. Note that in this case, and in contrast to vowel production, a simple homogeneous Dirichlet conditions is used for the acoustic pressure at the mouth exit. Future work will extend the presented approach to include outward sound wave propagation so as to account for radiation losses.

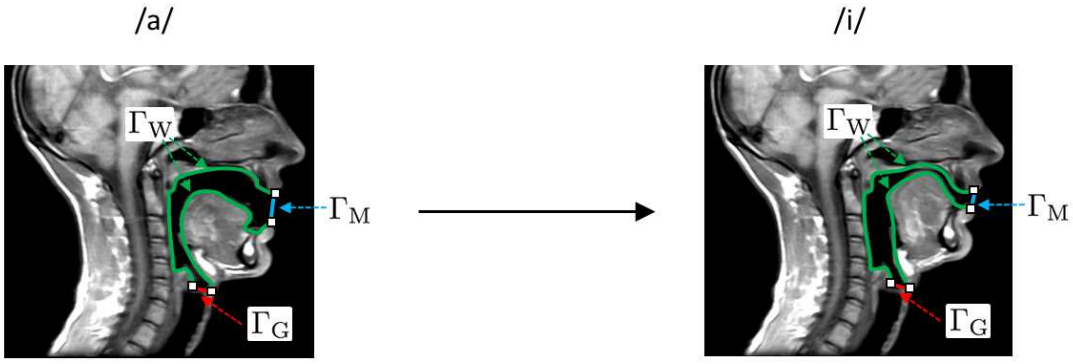


Fig. 8.1: A sketch of the computational domain Ω for vowels /a/ and /i/, from which diphthong /ai/ can be generated by transiting from one vowel to the other. The boundary regions used to define the boundary conditions in (8.8) are also indicated, with Γ_G standing for the glottal sectional area, Γ_W for the vocal tract walls and Γ_M for the mouth aperture.

8.3 Finite Element approach

8.3.1 Variational formulation

Equation (8.5), supplemented with the initial and boundary conditions in (8.8), constitutes the problem to be solved. Before expressing this equation in its variational form, let us introduce some additional notation and the functional framework without getting into much deep mathematical details. The integral between two arbitrary functions f and g in Ω will be written as

$$(f, g) := \int_{\Omega} f g d\Omega, \tag{8.9}$$

while integrals over boundaries will be explicitly indicated, for instance

$$(f, g)_{\Gamma_G} := \int_{\Gamma_G} f g d\Gamma_G. \tag{8.10}$$

The space of functions whose elements are square integrable in Ω will be denoted as usual by $L^2(\Omega)$, while $H^1(\Omega)$ will be used to denote the space of functions in $L^2(\Omega)$ that also have square integrable spatial derivatives. V_p and $\boldsymbol{V_u}$ will represent the space of functions that satisfy the Dirichlet boundary conditions of the problem and belong to $H^1(\Omega)$ for every time instant for the scalar pressure p and the vector velocity $\boldsymbol{u}$, respectively, whereas V_q and $\boldsymbol{V_v}$ will respectively stand for those functions in $H^1(\Omega)$ whose values vanish on the boundaries for the scalar test function q and the vector test function $\boldsymbol{v}$.

The weak formulation of (8.5) is found as usual by multiplying Eq. (8.5a) with a scalar test function q, Eq. (8.5b) with a vector test function $\boldsymbol{v}$, and integrating over the computational domain Ω. The variational problem can then be formulated as that of finding $p \in V_p$ and $\boldsymbol{u} \in \boldsymbol{V_u}$ such that

$$\frac{1}{\rho_0 c_0^2}(\partial_t p, q) - \frac{1}{\rho_0 c_0^2}(\boldsymbol{u}_{\text{dom}} \cdot \nabla p, q) + (\nabla \cdot \boldsymbol{u}, q) = (f_p, q), \tag{8.11a}$$

$$\rho_0(\partial_t \boldsymbol{u}, \boldsymbol{v}) - \rho_0(\boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}, \boldsymbol{v}) + (\nabla p, \boldsymbol{v}) = (\boldsymbol{f_u}, \boldsymbol{v}), \tag{8.11b}$$

for all $q \in V_q$ and $\boldsymbol{v} \in \boldsymbol{V_u}$. Next, integrating by parts the last term in the l.h.s of (8.11a), applying the divergence theorem and considering the boundary conditions in (8.8) we obtain

$$\frac{1}{\rho_0 c_0^2}(\partial_t p, q) - \frac{1}{\rho_0 c_0^2}(\boldsymbol{u}_{\text{dom}} \cdot \nabla p, q) - (\boldsymbol{u}, \nabla q) + (p/Z_w, q)_{\Gamma_{\text{W}}} = (f_p, q) - (g, q)_{\Gamma_{\text{G}}}, \tag{8.12a}$$

$$\rho_0(\partial_t \boldsymbol{u}, \boldsymbol{v}) - \rho_0(\boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}, \boldsymbol{v}) + (\nabla p, \boldsymbol{v}) = (\boldsymbol{f_u}, \boldsymbol{v}). \tag{8.12b}$$

Equation (8.12) corresponds to the ALE version of the third variational formulation for the wave equation in mixed form in Badia *et al.* (2014); Espinoza *et al.* (2014), which involves weak imposition of particle velocity boundary conditions, whilst strong imposition of the acoustic pressure at boundaries.

On the other hand, to shorten notation, equation (8.12) can be rewritten as

$$\mathcal{B}_1([p, \boldsymbol{u}], [q, \boldsymbol{v}]) = \mathcal{L}_1(q) \qquad \forall q_h \in V_{q,h}, \tag{8.13a}$$

$$\mathcal{B}_2([p, \boldsymbol{u}], [q, \boldsymbol{v}]) = \mathcal{L}_2(\boldsymbol{v}) \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_{\boldsymbol{u},h}, \tag{8.13b}$$

where we have defined the bilinear forms

$$\mathcal{B}_1([p, \boldsymbol{u}], [q, \boldsymbol{v}]) := \frac{1}{\rho_0 c_0^2}(\partial_t p, q) - \frac{1}{\rho_0 c_0^2}(\boldsymbol{u}_{\text{dom}} \cdot \nabla p, q) - (\boldsymbol{u}, \nabla q) + (p/Z_w, q)_{\Gamma_{\text{W}}}, \tag{8.14a}$$

$$\mathcal{B}_2([p, \boldsymbol{u}], [q, \boldsymbol{v}]) := \rho_0(\partial_t \boldsymbol{u}, \boldsymbol{v}) - \rho_0(\boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}, \boldsymbol{v}) + (\nabla p, \boldsymbol{v}), \tag{8.14b}$$

and the linear forms corresponding to the forcing terms by

$$\mathcal{L}_1(q) := (f_p, q) - (g, q)_{\Gamma_{\text{G}}}, \tag{8.15a}$$

$$\mathcal{L}_2(\boldsymbol{v}) := (\boldsymbol{f_u}, \boldsymbol{v}). \tag{8.15b}$$

8.3.2 Spatial discretization: subgrid stabilized scheme

The first step to find a numerical solution to Eq. (8.13) consists in the spatial discretization. Let us partition the computational domain Ω in n_{el} elements with n_p nodes. Given the finite dimensional spaces $V_{p,h} \subset V_p$ and $\boldsymbol{V}_{\boldsymbol{u},h} \subset \boldsymbol{V}_{\boldsymbol{u}}$, the Galerkin finite element approach to problem (8.13) consists in finding $p_h \in V_{p,h}$ and $\boldsymbol{u}_h \in \boldsymbol{V}_{\boldsymbol{u},h}$ such that

$$\mathcal{B}_1([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h]) = \mathcal{L}_1(q_h) \qquad \forall q_h \in V_{q,h}, \tag{8.16a}$$

$$\mathcal{B}_2([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h]) = \mathcal{L}_2(\boldsymbol{v}_h) \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_{\boldsymbol{u},h}, \tag{8.16b}$$

where the subindex h denotes the discrete version of the continuous variables. The standard Galerkin approach presented in (8.16) needs to satisfy an inf-sup compatibility constraint, which results in stability problems when equal interpolations for the approximated acoustic pressure and velocity fields are used. Next we introduce a variational subgrid scale stabilization strategy to circumvent this difficulty and thus to facilitate code implementation. This consists in splitting the exact solutions p and $\boldsymbol{u}$ into the finite element components p_h and $\boldsymbol{u}_h$ that can be resolved by the computational mesh, plus the subscales p' and $\boldsymbol{u}'$ which cannot be computed, and have to be somehow modeled. Substitution of $p = p_h + p'$ and $\boldsymbol{u} = \boldsymbol{u}_h + \boldsymbol{u}'$ into (8.12) yields two sets of equations. The first one accounting for the large solvable scales is given by

$$\begin{aligned}\frac{1}{\rho_0 c_0^2}\big[(\partial_t p_h, q_h) + (\partial_t p', q_h)\big] - \frac{1}{\rho_0 c_0^2}\big[(\boldsymbol{u}_{\text{dom}} \cdot \nabla p_h, q_h) + (\boldsymbol{u}_{\text{dom}} \cdot \nabla p', q_h)\big] \\ - \big[(\boldsymbol{u}_h, \nabla q_h) + (\boldsymbol{u}', \nabla q_h)\big] + (p_h/Z_w, q_h)_{\Gamma_{\text{W}}} = (f_p, q_h) - (g, q_h)_{\Gamma_{\text{G}}},\end{aligned} \tag{8.17a}$$

$$\begin{aligned}\rho_0\big[(\partial_t \boldsymbol{u}_h, \boldsymbol{v}_h) + (\partial_t \boldsymbol{u}', \boldsymbol{v}_h)\big] - \rho_0\big[(\boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h) + (\boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}', \boldsymbol{v}_h)\big] \\ + \big[(\nabla p_h, \boldsymbol{v}_h) + (\nabla p', \boldsymbol{v}_h)\big] = (\boldsymbol{f}_{\boldsymbol{u}}, \boldsymbol{v}_h),\end{aligned} \tag{8.17b}$$

where we have imposed in the boundary integrals that the subscales vanish on the element contours. Assuming quasi-static subscales (i.e., $\partial_t p' \approx 0$, $\partial_t \boldsymbol{u}' \approx 0$) and using the linear and bilinear forms defined in (8.14) and (8.15), Eq. (8.17) can be rearranged as

$$\mathcal{B}_1([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h]) - \frac{1}{\rho_0 c_0^2}(\boldsymbol{u}_{\text{dom}} \cdot \nabla p', q_h) - (\boldsymbol{u}', \nabla q_h) = \mathcal{L}_1(q_h), \tag{8.18a}$$

$$\mathcal{B}_2([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h]) - \rho_0(\boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}', \boldsymbol{v}_h) + (\nabla p', \boldsymbol{v}_h) = \mathcal{L}_2(\boldsymbol{v}_h). \tag{8.18b}$$

Integrating by parts yields

$$\begin{aligned}\mathcal{B}_1([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h]) + \frac{1}{\rho_0 c_0^2}(p', \boldsymbol{u}_{\text{dom}} \cdot \nabla q_h) + \frac{1}{\rho_0 c_0^2}(p', \nabla \cdot \boldsymbol{u}_{\text{dom}}\, q_h) \\ - (\boldsymbol{u}', \nabla q_h) = \mathcal{L}_1(q_h),\end{aligned} \tag{8.19a}$$

$$\begin{aligned}\mathcal{B}_2([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h]) + \rho_0(\boldsymbol{u}', \boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{v}_h) + \rho_0(\boldsymbol{u}', \nabla \cdot \boldsymbol{u}_{\text{dom}}\, \boldsymbol{v}_h) \\ - (p', \nabla \cdot \boldsymbol{v}_h) = \mathcal{L}_2(\boldsymbol{v}_h).\end{aligned} \tag{8.19b}$$

The terms $(p', \nabla\cdot\boldsymbol{u}_{\text{dom}}\, q_h)$ in (8.19a) and $(\boldsymbol{u}', \nabla\cdot\boldsymbol{u}_{\text{dom}}\, \boldsymbol{v}_h)$ in (8.19b) will be disregarded for simplicity. They do not affect the consistency of the proposed approach and the remaining terms will be shown to produce enough stability for the problem at hand.

Let us now focus on the second equation coming from splitting the unknowns in $p_h + p'$ and $\boldsymbol{u}_h + \boldsymbol{u}'$. This equation, governing the behavior of the subscales, is given by

$$\frac{1}{\rho_0 c_0^2}\big[(\partial_t p_h, q') + (\partial_t p', q')\big] - \frac{1}{\rho_0 c_0^2}\big[(\boldsymbol{u}_{\text{dom}}\cdot\nabla p_h, q') + (\boldsymbol{u}_{\text{dom}}\cdot\nabla p', q')\big] + \big[(\nabla\cdot\boldsymbol{u}_h, q') + (\nabla\cdot\boldsymbol{u}', q')\big] = (f_p, q'), \tag{8.20a}$$

$$\rho_0\big[(\partial_t \boldsymbol{u}_h, \boldsymbol{v}') + (\partial_t \boldsymbol{u}', \boldsymbol{v}')\big] - \rho_0\big[(\boldsymbol{u}_{\text{dom}}\cdot\nabla \boldsymbol{u}_h, \boldsymbol{v}') + (\boldsymbol{u}_{\text{dom}}\cdot\nabla \boldsymbol{u}', \boldsymbol{v}')\big] + \big[(\nabla p_h, \boldsymbol{v}') + (\nabla p', \boldsymbol{v}')\big] = (\boldsymbol{f}_{\boldsymbol{u}}, \boldsymbol{v}'), \tag{8.20b}$$

where integration by parts has been applied to $(\boldsymbol{u}_h, \nabla q')$ and $(\boldsymbol{u}', \nabla q')$ to obtain the last two terms in (8.20a), and the boundary integrals have vanished given that subscales become zero at the element contours. Note then that (8.20) is nothing but the L^2-projection of each term onto the space of subscales. Denoting with $\mathcal{P}'_p$ and $\mathcal{P}'_{\boldsymbol{u}}$ these projections and assuming again quasi-static subscales, (8.20) can be rewritten as

$$\mathcal{P}'_p\Big(-\frac{1}{\rho_0 c_0^2}\boldsymbol{u}_{\text{dom}}\cdot\nabla p' + \nabla\cdot\boldsymbol{u}'\Big) = R_{p,h}, \tag{8.21a}$$

$$\mathcal{P}'_{\boldsymbol{u}}\Big(-\rho_0\boldsymbol{u}_{\text{dom}}\cdot\nabla\boldsymbol{u}' + \nabla p'\Big) = \boldsymbol{R}_{\boldsymbol{u},h}, \tag{8.21b}$$

where we have defined the residual functions $R_{p,h}$ and $\boldsymbol{R}_{\boldsymbol{u},h}$ of the finite element approximation onto the subscale space as

$$R_{p,h} := \mathcal{P}'_p\left[f_p - \left(\frac{1}{\rho_0 c_0^2}\partial_t p_h - \frac{1}{\rho_0 c_0^2}\boldsymbol{u}_{\text{dom}}\cdot\nabla p_h + \nabla\cdot\boldsymbol{u}_h\right)\right], \tag{8.22a}$$

$$\boldsymbol{R}_{\boldsymbol{u},h} := \mathcal{P}'_{\boldsymbol{u}}\Big[\boldsymbol{f}_{\boldsymbol{u}} - \big(\rho_0\partial_t\boldsymbol{u}_h - \rho_0\boldsymbol{u}_{\text{dom}}\cdot\nabla\boldsymbol{u}_h + \nabla p_h\big)\Big]. \tag{8.22b}$$

As said, (8.21) is an equation for the subscales p' a $\boldsymbol{u}'$ whose solution is unknown and has to be modeled. A reasonable approximation is that of taking (Codina, 2008)

$$p' \simeq \tau_p R_{p,h}, \tag{8.23a}$$

$$\boldsymbol{u}' \simeq \tau_{\boldsymbol{u}}\boldsymbol{R}_{\boldsymbol{u},h}, \tag{8.23b}$$

with τ_p and $\tau_{\boldsymbol{u}}$ standing for scalar stabilization parameters to be determined. For simplicity, in this case we have already disregarded the cross influence of the velocity subscales components in (8.23b) and also their impact in the equation for the pressure subscales (8.23a) and viceversa, as will be latter done when designing the stabilization parameters in Section 8.3.3. Note also that we have assumed the same value $\tau_{\boldsymbol{u}}$ for each

component of the velocity subscale vector. Next, inserting (8.23) in the equation for the large scales we get

$$\mathcal{B}_1([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h]) + \frac{1}{\rho_0 c_0^2}(\tau_p R_{p,h}, \boldsymbol{u}_{\text{dom}} \cdot \nabla q_h) - (\tau_{\boldsymbol{u}} \boldsymbol{R}_{\boldsymbol{u},h}, \nabla q_h) = \mathcal{L}_1(q_h), \tag{8.24a}$$

$$\mathcal{B}_2([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h]) + \rho_0(\tau_{\boldsymbol{u}} \boldsymbol{R}_{\boldsymbol{u},h}, \boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{v}_h) - (\tau_p R_{p,h}, \nabla \cdot \boldsymbol{v}_h) = \mathcal{L}_2(\boldsymbol{v}_h). \tag{8.24b}$$

Note in Eq. (8.24) that the bilinear forms $\mathcal{B}_1([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h])$ and $\mathcal{B}_2([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h])$ correspond to the Galerkin approach (8.16) and that the remaining terms in the l.h.s. are the additional stabilization terms that facilitate using equal interpolations for the acoustic pressure and acoustic particle velocity.

The last step now consists in approximating the projection operators appearing in the residual functions in (8.22). To do so, there exist different options in the literature. In the so called OSS (Orthogonal Subgrid Scale) method (Codina, 2002; Codina *et al.*, 2007), the subscales are assumed to lay in a space orthogonal to the finite element space so that $\mathcal{P}'$ can be computed as

$$\mathcal{P}' = \boldsymbol{I} - \Pi_h, \tag{8.25}$$

Π_h being the L^2-projection onto the finite element space, and $\boldsymbol{I}$ the identity matrix. Alternatively, in the more classical ASGS (Algebraic Subgrid Scale) approach we simply take

$$\mathcal{P}' = \boldsymbol{I} \tag{8.26}$$

over the space of finite element residuals. The latter has been the option adopted in this work, which yields from (8.24) the expression

$$\begin{aligned}
&\mathcal{B}_1([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h]) \\
&\quad + \frac{1}{\rho_0 c_0^2}\left(\tau_p\left[f_p - \frac{1}{\rho_0 c_0^2}\partial_t p_h + \frac{1}{\rho_0 c_0^2}\boldsymbol{u}_{\text{dom}} \cdot \nabla p_h - \nabla \cdot \boldsymbol{u}_h\right], \boldsymbol{u}_{\text{dom}} \cdot \nabla q_h\right) \\
&\quad - \Big(\tau_{\boldsymbol{u}}\big[\boldsymbol{f}_{\boldsymbol{u}} - \rho_0 \partial_t \boldsymbol{u}_h + \rho_0 \boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}_h - \nabla p_h\big], \nabla q_h\Big) \\
&\quad = \mathcal{L}_1(q_h),
\end{aligned} \tag{8.27a}$$

$$\begin{aligned}
&\mathcal{B}_2([p_h, \boldsymbol{u}_h], [q_h, \boldsymbol{v}_h]) \\
&\quad + \rho_0\Big(\tau_{\boldsymbol{u}}\big[\boldsymbol{f}_{\boldsymbol{u}} - \rho_0 \partial_t \boldsymbol{u}_h + \rho_0 \boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}_h - \nabla p_h\big], \boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{v}_h\Big) \\
&\quad - \left(\tau_p\left[f_p - \frac{1}{\rho_0 c_0^2}\partial_t p_h + \frac{1}{\rho_0 c_0^2}\boldsymbol{u}_{\text{dom}} \cdot \nabla p_h - \nabla \cdot \boldsymbol{u}_h\right], \nabla \cdot \boldsymbol{v}_h\right) \\
&\quad = \mathcal{L}_2(\boldsymbol{v}_h).
\end{aligned} \tag{8.27b}$$

8.3.3 Design of the stabilization parameter

In this section we will derive expressions for the stabilization parameters τ_p and $\tau_{\boldsymbol{u}}$. Let us first rewrite the equation for the subscales (8.21) in compact matrix form. Assuming the summation convention over repeated indexes this becomes

$$\mathcal{P}'(\boldsymbol{A}_i\partial_i\boldsymbol{U}') = \mathcal{P}'[\boldsymbol{F} - (\partial_t\boldsymbol{U}_h + \boldsymbol{A}_i\partial_i\boldsymbol{U}_h)] =: \boldsymbol{R}_h, \tag{8.28}$$

where $\boldsymbol{U} = \boldsymbol{U}_h + \boldsymbol{U}' = [p_h, \boldsymbol{u}_h]^\top + [p', \boldsymbol{u}']^\top$ is the vector of unknowns, $\boldsymbol{F} = [f_p, \boldsymbol{f}_{\boldsymbol{u}}]^\top$ is the vector with the external forcing terms, $\boldsymbol{R}_h = [R_{p,h}, \boldsymbol{R}_{\boldsymbol{u},h}]^\top$ is the residual of the finite element approximation onto the subscale space, $\mathcal{P}'$ is the L^2-projection onto the subscales, and the matrices $\boldsymbol{A}_i$ are defined by

$$\boldsymbol{A}_1 = \begin{pmatrix} -\mu_p u_{d1} & 1 & 0 & 0 \\ 1 & -\mu_{\boldsymbol{u}} u_{d1} & 0 & 0 \\ 0 & 0 & -\mu_{\boldsymbol{u}} u_{d1} & 0 \\ 0 & 0 & 0 & -\mu_{\boldsymbol{u}} u_{d1} \end{pmatrix}, \quad \boldsymbol{A}_2 = \begin{pmatrix} -\mu_p u_{d2} & 0 & 1 & 0 \\ 0 & -\mu_{\boldsymbol{u}} u_{d2} & 0 & 0 \\ 1 & 0 & -\mu_{\boldsymbol{u}} u_{d2} & 0 \\ 0 & 0 & 0 & -\mu_{\boldsymbol{u}} u_{d2} \end{pmatrix},$$

$$\boldsymbol{A}_3 = \begin{pmatrix} -\mu_p u_{d3} & 0 & 0 & 1 \\ 0 & -\mu_{\boldsymbol{u}} u_{d3} & 0 & 0 \\ 0 & 0 & -\mu_{\boldsymbol{u}} u_{d3} & 0 \\ 1 & 0 & 0 & -\mu_{\boldsymbol{u}} u_{d3} \end{pmatrix}. \tag{8.29}$$

To ease the notation and to facilitate comparisons with the work in Codina (2008) as well, we have introduced the parameters $\mu_p \equiv (\rho_0 c_0^2)^{-1}$ and $\mu_{\boldsymbol{u}} \equiv \rho_0$ which fulfill $c_0 = (\mu_p \mu_{\boldsymbol{u}})^{-1/2}$. u_{di} denote the components of the mesh velocity vector at each point, $\boldsymbol{u}_{\text{dom}} = (u_{d1}, u_{d2}, u_{d3})^\top$. Analogously, defining $\boldsymbol{\tau} = \text{diag}(\tau_p, \tau_{\boldsymbol{u}}, \tau_{\boldsymbol{u}}, \tau_{\boldsymbol{u}})$, the approximation for the subscales (8.23) can be rewritten as

$$\boldsymbol{U}' = \boldsymbol{\tau}\boldsymbol{R}_h. \tag{8.30}$$

The procedure to find the stabilization matrix $\boldsymbol{\tau}$, to be described below, involves dealing with the norms of the residual vector $\boldsymbol{R}_h$, which has dimensions of force, and of the subscale velocity vector $\boldsymbol{U}'$. However, we note that contrary to $\boldsymbol{U}^\top\boldsymbol{F}$, products like $\boldsymbol{F}^\top\boldsymbol{F}$ or $\boldsymbol{U}^\top\boldsymbol{U}$ are not dimensionally well defined. For instance, $\boldsymbol{F}^\top\boldsymbol{F} = f_p^2 + f_{u_1}^2 + f_{u_2}^2 + f_{u_3}^2$ but $[f_p^2] \neq [f_{u_i}^2]$. Analogously $\boldsymbol{U}^\top\boldsymbol{U} = p^2 + u_1^2 + u_2^2 + u_3^2$ and again, obviously, $[p^2] \neq [u_i^2]$. To correctly define these products we can introduce the weighting matrix $\boldsymbol{M}$ (Codina, 2008),

$$\boldsymbol{M} = \begin{pmatrix} m_p & 0 & 0 & 0 \\ 0 & m_{\boldsymbol{u}} & 0 & 0 \\ 0 & 0 & m_{\boldsymbol{u}} & 0 \\ 0 & 0 & 0 & m_{\boldsymbol{u}} \end{pmatrix}, \quad m_p := \sqrt{\mu_{\boldsymbol{u}}/\mu_p} = \rho_0 c_0, \quad m_{\boldsymbol{u}} := \sqrt{\mu_p/\mu_{\boldsymbol{u}}} = \frac{1}{\rho_0 c_0}, \tag{8.31}$$

and define the norms

$$|\boldsymbol{F}|^2_{\boldsymbol{M}} := \boldsymbol{F}^\top \boldsymbol{M}\boldsymbol{F}, \tag{8.32a}$$

$$|\boldsymbol{U}|^2_{\boldsymbol{M}^{-1}} := \boldsymbol{U}^\top \boldsymbol{M}^{-1}\boldsymbol{U}. \tag{8.32b}$$

We can readily check that $|\boldsymbol{F}|^2_{\boldsymbol{M}}$ is now dimensionally correct because $[m_p f_p^2] = [m_{\boldsymbol{u}} f_{u_i}^2] = ML^{-2}T^{-3}$, with M, L and T respectively standing for dimensions of mass, length and time. With regard to the unknowns $\boldsymbol{U}$, we have to use the inverse of the $\boldsymbol{M}$ matrix to get the right dimensions. Indeed, it follows that $|\boldsymbol{U}|^2_{\boldsymbol{M}^{-1}}$ is well defined because $[m_p^{-1}p^2] = [m_{\boldsymbol{u}}^{-1}u_i^2] = MT^{-3}$. Thus we can use $\boldsymbol{M}$ and $\boldsymbol{M}^{-1}$ to respectively define products between any two force vectors $\boldsymbol{F}_1^\top \boldsymbol{M}\boldsymbol{F}_2$ and unknowns $\boldsymbol{U}_1^\top \boldsymbol{M}^{-1}\boldsymbol{U}_2$. We can proceed similarly and define a $\boldsymbol{M}$-weighted product between two arbitrary matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ as $\boldsymbol{B}_1^\top \boldsymbol{M}\boldsymbol{B}_2$. The squared $\boldsymbol{M}$-pointwise norm of a matrix $\boldsymbol{B}$ will be given by

$$|\boldsymbol{B}|^2_{\boldsymbol{M}} := \sup_{|\boldsymbol{X}|_{\boldsymbol{M}^{-1}}=1} \boldsymbol{X}^\top(\boldsymbol{B}^\top \boldsymbol{M}\boldsymbol{B})\boldsymbol{X}. \tag{8.33}$$

Equipped with the weighting matrix $\boldsymbol{M}$ and for regular enough force vector functions, we can also define the following vector function scalar products,

$$(\boldsymbol{U}, \boldsymbol{F}) = \int_\Omega \boldsymbol{U}^\top \boldsymbol{F} d\Omega, \tag{8.34a}$$

$$(\boldsymbol{F}_1, \boldsymbol{F}_2)_{\boldsymbol{M}} := \int_\Omega \boldsymbol{F}_1^\top \boldsymbol{M}\boldsymbol{F}_2 d\Omega, \tag{8.34b}$$

$$(\boldsymbol{U}_1, \boldsymbol{U}_2)_{\boldsymbol{M}^{-1}} := \int_\Omega \boldsymbol{U}_1^\top \boldsymbol{M}^{-1}\boldsymbol{U}_2 d\Omega, \tag{8.34c}$$

whose induced norms will be denoted by $\|\cdot\|$, $\|\cdot\|_{\boldsymbol{M}}$ and $\|\cdot\|_{\boldsymbol{M}^{-1}}$.

A suitable option to find the matrix of stabilization parameters $\boldsymbol{\tau}$ is based on a Fourier analysis of the equation for the subscales (8.28). Let use a hat symbol to denote Fourier transformed functions. Transforming (8.28) to the wave number domain allows us to bound the norm of the Fourier transformed residual $\hat{\boldsymbol{R}}_h$ as (Codina, 2008; Codina *et al.*, 2008)

$$\|\hat{\boldsymbol{R}}_h\|^2_{\boldsymbol{M}} \leq \left| -i\frac{1}{h} k_j^0 \boldsymbol{A}_j \right|^2_{\boldsymbol{M}} \|\hat{\boldsymbol{U}}'\|^2_{\boldsymbol{M}^{-1}}, \tag{8.35}$$

with k_j^0 standing for particular dimensionless wave number components that arise from the application of the mean value theorem and that, in practice, will be treated as algorithmic constants (see below). On the other hand, from the Fourier transform of the approximation for the subscales (8.30), we can get the bound

$$\|\hat{\boldsymbol{R}}_h\|^2_{\boldsymbol{M}} \leq |\boldsymbol{\tau}^{-1}|^2_{\boldsymbol{M}} \, \|\hat{\boldsymbol{U}}'\|^2_{\boldsymbol{M}^{-1}}. \tag{8.36}$$

Therefore, (8.35) and (8.36) provide a means to get the stabilization matrix $\boldsymbol{\tau}$ by equating

$$|\boldsymbol{\tau}^{-1}|^2_{\boldsymbol{M}} = \left| -i\frac{1}{h}k_j^0 \boldsymbol{A}_j \right|^2_{\boldsymbol{M}}. \tag{8.37}$$

Using the definition of the squared $\boldsymbol{M}$-pointwise norm of a matrix (8.33) we get

$$\sup_{|\boldsymbol{X}|_{\boldsymbol{M}^{-1}}=1} \boldsymbol{X}^\top(\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1})\boldsymbol{X} = \sup_{|\boldsymbol{X}|_{\boldsymbol{M}^{-1}}=1} \boldsymbol{X}^\top(\frac{1}{h^2}k_i^0 k_j^0 \boldsymbol{A}_i^\top \boldsymbol{M}\boldsymbol{A}_j)\boldsymbol{X}. \tag{8.38}$$

In practice, we can implement the above condition by computing the spectra of matrices $\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1}$ and $\frac{1}{h^2}k_i^0 k_j^0 \boldsymbol{A}_i^\top \boldsymbol{M}\boldsymbol{A}_j$ with respect to matrix $\boldsymbol{M}^{-1}$, and then imposing their spectral radii to be the same.

Taking into account the definitions for $\boldsymbol{A}_i$ in (8.29) and for the scaling matrix $\boldsymbol{M}$ in (8.31), we get for the problem at hand,

$$\frac{1}{h^2}k_i^0 k_j^0 \boldsymbol{A}_i^\top \boldsymbol{M}\boldsymbol{A}_j = \frac{1}{h^2}\times$$

$$\begin{pmatrix}
\begin{matrix} m_{\boldsymbol{u}}|\boldsymbol{k}^0|^2 \\ +\beta_p|\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}|^2 \end{matrix} & -\alpha k_1^0(\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}) & -\alpha k_2^0(\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}) & -\alpha k_3^0(\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}) \\
-\alpha k_1^0(\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}) & \begin{matrix} m_p(k_1^0)^2 \\ +\beta_{\boldsymbol{u}}|\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}|^2 \end{matrix} & k_1^0k_2^0m_p & k_1^0k_3^0m_p \\
-\alpha k_2^0(\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}) & k_1^0k_2^0m_p & \begin{matrix} m_p(k_2^0)^2 \\ +\beta_{\boldsymbol{u}}|\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}|^2 \end{matrix} & k_2^0k_3^0m_p \\
-\alpha k_3^0(\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}) & k_1^0k_3^0m_p & k_2^0k_3^0m_p & \begin{matrix} m_p(k_3^0)^2 \\ +\beta_{\boldsymbol{u}}|\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}|^2 \end{matrix}
\end{pmatrix} \tag{8.39a}$$

with

$$\alpha \equiv (m_p\mu_p + m_{\boldsymbol{u}}\mu_{\boldsymbol{u}}), \quad \beta_p \equiv m_p\mu_p^2, \quad \beta_{\boldsymbol{u}} \equiv m_{\boldsymbol{u}}\mu_{\boldsymbol{u}}^2. \tag{8.39b}$$

Let us denote the spectrum of the generalized eigenvalue problem $\boldsymbol{B}\boldsymbol{U} = \lambda\boldsymbol{M}^{-1}\boldsymbol{U}$, λ being an eigenvalue, by $\text{Spec}_{\boldsymbol{M}^{-1}}(\boldsymbol{B})$, and its spectral radius by $\rho(\boldsymbol{B})$. The spectrum of $k_i^0k_j^0\boldsymbol{A}_i^\top\boldsymbol{M}\boldsymbol{A}_j$ can be analytically worked out and is given by

$$\text{Spec}_{\boldsymbol{M}^{-1}}(k_i^0k_j^0\boldsymbol{A}_i^\top\boldsymbol{M}\boldsymbol{A}_j) = \left\{ \left(\frac{\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}}{c_0} + |\boldsymbol{k}^0|\right)^2, \left(\frac{\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}}{c_0}\right)^2, \left(\frac{\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}}{c_0}\right)^2, \left(\frac{\boldsymbol{k}^0\cdot\boldsymbol{u}_{\text{dom}}}{c_0} - |\boldsymbol{k}^0|\right)^2 \right\}, \tag{8.40}$$

where, as seen, the eigenvalue $c_0^{-2}(\boldsymbol{k}^0 \cdot \boldsymbol{u}_{\text{dom}})^2$ has multiplicity 2. As observed, the spectral radius of $k_i^0 k_j^0 \boldsymbol{A}_i^\top \boldsymbol{M} \boldsymbol{A}_j$ is given by

$$\rho(k_i^0 k_j^0 \boldsymbol{A}_i^\top \boldsymbol{M} \boldsymbol{A}_j) = \left(\frac{\boldsymbol{k}^0 \cdot \boldsymbol{u}_{\text{dom}}}{c_0} + |\boldsymbol{k}^0| \right)^2 \equiv \left(\frac{C_1 |\boldsymbol{u}_{\text{dom}}|}{c_0} + C_2 \right)^2. \tag{8.41}$$

Since $\boldsymbol{k}^0$ is an unknown wave number, in the r.h.s equivalence we have taken

$$\boldsymbol{k}^0 \cdot \boldsymbol{u}_{\text{dom}} = |\boldsymbol{k}^0||\boldsymbol{u}_{\text{dom}}| \cos\theta \equiv C_1 |\boldsymbol{u}_{\text{dom}}|, \qquad |\boldsymbol{k}^0| \equiv C_2, \tag{8.42}$$

C_1, C_2 being constants that should be determined from numerical experiments. Taking the stabilization matrix $\boldsymbol{\tau}$ to be diagonal for simplicity, it follows that

$$\text{Spec}_{\boldsymbol{M}^{-1}}(\boldsymbol{\tau}^{-1} \boldsymbol{M} \boldsymbol{\tau}^{-1}) = \left\{ \left(\frac{m_p}{\tau_p} \right)^2, \left(\frac{m_{\boldsymbol{u}}}{\tau_{\boldsymbol{u}}} \right)^2, \left(\frac{m_{\boldsymbol{u}}}{\tau_{\boldsymbol{u}}} \right)^2, \left(\frac{m_{\boldsymbol{u}}}{\tau_{\boldsymbol{u}}} \right)^2 \right\}. \tag{8.43}$$

As we want to fulfill the condition

$$\rho(\boldsymbol{\tau}^{-1} \boldsymbol{M} \boldsymbol{\tau}^{-1}) = \rho(h^{-2} k_i^0 k_j^0 \boldsymbol{A}_i^\top \boldsymbol{M} \boldsymbol{A}_j), \tag{8.44}$$

we may achieve that by forcing every eigenvalue of $\boldsymbol{\tau}^{-1} \boldsymbol{M} \boldsymbol{\tau}^{-1}$ in (8.43) to equal to (8.41). This results in the following matrix of stabilization parameters

$$\boldsymbol{\tau} = \begin{pmatrix} \tau_p & 0 & 0 & 0 \\ 0 & \tau_{\boldsymbol{u}} & 0 & 0 \\ 0 & 0 & \tau_{\boldsymbol{u}} & 0 \\ 0 & 0 & 0 & \tau_{\boldsymbol{u}} \end{pmatrix} \quad \text{with,}$$

$$\begin{aligned} \tau_p &= \frac{c_0 h \sqrt{\mu_{\boldsymbol{u}}/\mu_p}}{C_1 |\boldsymbol{u}_{\text{dom}}| + c_0 C_2} = \frac{\rho_0 c_0^2 h}{C_1 |\boldsymbol{u}_{\text{dom}}| + c_0 C_2}, \\ \tau_{\boldsymbol{u}} &= \frac{c_0 h \sqrt{\mu_p/\mu_{\boldsymbol{u}}}}{C_1 |\boldsymbol{u}_{\text{dom}}| + c_0 C_2} = \frac{h}{C_1 \rho_0 |\boldsymbol{u}_{\text{dom}}| + \rho_0 c_0 C_2}, \end{aligned} \tag{8.45}$$

and where use has been made of the expressions for m_p and $m_{\boldsymbol{u}}$ in (8.31). In the case of a static domain $\boldsymbol{u}_{\text{dom}} = \boldsymbol{0}$ and we recover from (8.45) the stabilization parameters for the standard wave equation in mixed form (Codina, 2008),

$$\tau_p = Ch\sqrt{\mu_{\boldsymbol{u}}/\mu_p} \qquad \tau_{\boldsymbol{u}} = Ch\sqrt{\mu_p/\mu_{\boldsymbol{u}}}, \tag{8.46}$$

with C standing for a constant.

8.3.4 Time discretization: final numerical scheme

Up to now we have left the time continuous in the previous developments. As for the time discretization, we equally split the time interval $[0\ T]$ into N steps

$$0 < t^1 < t^2 < \cdots < t^{n-1} < t^n < t^{n+1} < \cdots < t^N \equiv T \tag{8.47}$$

such that $\Delta t := t^{n+1} - t^n$ is the time step size. For any time dependent function $g(t)$, g^n will stand for its evaluation at $t^n = n\Delta t$. The time evolution of (8.27) is then approximated using a second order backward differentiation formula (BDF2). Let us identify

$$\delta_t g^{n+1} := \frac{1}{2\Delta t}\left(3g^{n+1} - 4g^n + g^{n-1}\right), \tag{8.48}$$

so that the time discrete version of (8.27) becomes

$$\begin{aligned}
&\frac{1}{\rho_0 c_0^2}(\delta_t p_h^{n+1}, q_h) - \frac{1}{\rho_0 c_0^2}(\boldsymbol{u}_{\text{dom}} \cdot \nabla p_h^{n+1}, q_h) - (\boldsymbol{u}_h^{n+1}, \nabla q_h) + (p^{n+1}/Z_w, q_h)_{\Gamma_{\text{W}}} \\
&+ \frac{1}{\rho_0 c_0^2}\left(\tau_p\left[f_p^{n+1} - \frac{1}{\rho_0 c_0^2}\delta_t p_h^{n+1} + \frac{1}{\rho_0 c_0^2}\boldsymbol{u}_{\text{dom}} \cdot \nabla p_h^{n+1} - \nabla \cdot \boldsymbol{u}_h^{n+1}\right], \boldsymbol{u}_{\text{dom}} \cdot \nabla q_h\right) \\
&- \left(\tau_{\boldsymbol{u}}\left[\boldsymbol{f}_{\boldsymbol{u}}^{n+1} - \rho_0 \delta_t \boldsymbol{u}_h^{n+1} + \rho_0 \boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}_h^{n+1} - \nabla p_h^{n+1}\right], \nabla q_h\right) \\
&= (f_p^{n+1}, q_h) - (g^{n+1}, q_h)_{\Gamma_{\text{G}}},
\end{aligned} \tag{8.49a}$$

$$\begin{aligned}
&\rho_0(\delta_t \boldsymbol{u}_h^{n+1}, \boldsymbol{v}_h) - \rho_0(\boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}_h^{n+1}, \boldsymbol{v}_h) + (\nabla p_h^{n+1}, \boldsymbol{v}_h) \\
&+ \rho_0\left(\tau_{\boldsymbol{u}}\left[\boldsymbol{f}_{\boldsymbol{u}}^{n+1} - \rho_0 \delta_t \boldsymbol{u}_h^{n+1} + \rho_0 \boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}_h^{n+1} - \nabla p_h^{n+1}\right], \boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{v}_h\right) \\
&- \left(\tau_p\left[f_p^{n+1} - \frac{1}{\rho_0 c_0^2}\delta_t p_h^{n+1} + \frac{1}{\rho_0 c_0^2}\boldsymbol{u}_{\text{dom}} \cdot \nabla p_h^{n+1} - \nabla \cdot \boldsymbol{u}_h^{n+1}\right], \nabla \cdot \boldsymbol{v}_h\right) \\
&= (\boldsymbol{f}_{\boldsymbol{u}}^{n+1}, \boldsymbol{v}_h).
\end{aligned} \tag{8.49b}$$

Equation (8.49), with the stabilization parameters in (8.45), is the final fully discrete scheme in space and time proposed in this work. Note that in contrast to the derived numerical scheme (2.18) for vowel production (see Section 2.3.3), this one corresponds to an implicit scheme, which requires matrix inversion (higher computational cost) but provides a less restrictive Courant-Friedrich-Levy condition compared to explicit schemes, allowing us to use longer time steps. For the particular case of voice production, $f_p = 0$ and $\boldsymbol{f}_{\boldsymbol{u}} = \boldsymbol{0}$. They have been included in the development for completeness.

8.4 General description of the FE simulations

In Fig. 8.2 a block diagram that describes the methodology used to perform a FE simulation of a diphthong sound is represented. One can identify the same blocks of the computational voice engine presented in Fig. 1.1 of Chapter 1:

- **Vocal tract geometry model:** this model generates the set of static vocal tracts for different vowel sounds and interpolate their boundary values to produce a diphthong transition. This model also provides an initial FE mesh.

- **Glottal source model:** this model generates a train of glottal pulses produced by the vocal cords to be introduced at the glottal cross-section of the vocal tract.

- **Vocal tract acoustic model:** using as inputs the initial FE mesh and the boundary node coordinates provided by the geometry model, this model first obtain a new FE mesh for every time instant by computing the position of the inner nodes. The mesh velocity $\boldsymbol{u}_{\text{dom}}$ is then derived using the current and previous vocal tract meshes. Finally, using the glottal pulses generated by the glottal model, the ALE mixed wave equation is resolved for each mesh. The acoustic pressure is tracked in a node close to the mouth aperture and transformed to an audio file.

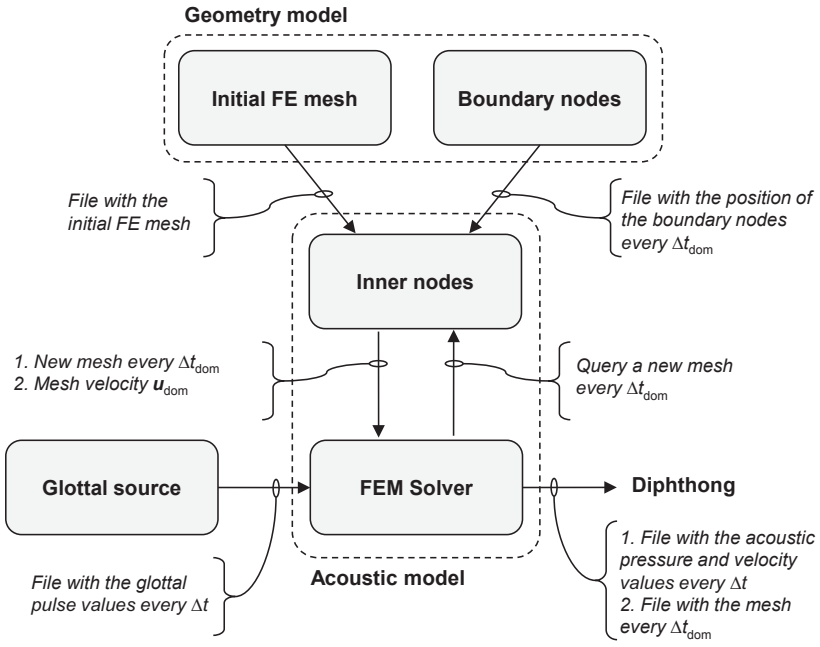


Fig. 8.2: Block diagram for the FEM simulation of diphthongs.

Due to CFL restrictions, the use of a very small time step for the acoustic wave equation is usually required, whereas it might not be necessary the same small resolution to capture the vocal tract movement. A different time step can be used for the acoustic wave equation (Δt) and for updating the mesh (Δt_{dom}), of course fulfilling $\Delta t \leq \Delta t_{\text{dom}}$. Logically, the mesh velocity $\boldsymbol{u}_{\text{dom}}$ will be zero at those time instants where no new mesh is generated.

In what concerns the motion of the computational mesh, it is driven by prescribed displacements on the boundary nodes generated by the geometry model, according to the vocal tract profiles that represent the transition from one vowel to another. The boundary motion is smoothly transmitted to the inner mesh nodes through diffusion, i.e. by solving the Laplacian equation for the node displacements $\boldsymbol{w}(\boldsymbol{x}, t)$. Therefore, every Δt_{dom} we solve using FEM the additional equation

$$\nabla^2 \boldsymbol{w}^{n+1} = 0 \qquad \text{in } \Omega,\ t = t^{n+1}, \tag{8.50a}$$

with boundary conditions

$$\boldsymbol{w}^{n+1} = \boldsymbol{x}^{n+1}_{\text{walls}} - \boldsymbol{x}^{n}_{\text{walls}} \qquad \text{on } \Gamma_{\text{W}},\ t = t^{n+1}, \tag{8.50b}$$

$$\boldsymbol{w}^{n+1} \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\text{G}},\ t = t^{n+1}, \tag{8.50c}$$

$$\boldsymbol{w}^{n+1} \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\text{M}},\ t = t^{n+1}, \tag{8.50d}$$

the mesh node positions being updated according to

$$\boldsymbol{x}^{n+1} = \boldsymbol{x}^{n} + \boldsymbol{w}^{n+1}. \tag{8.50e}$$

Note that in addition to the prescribed vocal tract wall displacement in (8.50b), two additional boundary conditions are imposed on the glottal (8.50c) and mouth (8.50d) cross-sections. This limits any movement of the nodes in the tangential direction, which would artificially lengthen the vocal tract. The velocity of the computational mesh appearing in Eq. (8.5) is simply computed at time step $n+1$ for a node i, with coordinates $\boldsymbol{x}_i(t)$, as

$$\boldsymbol{u}^{n+1}_{\text{dom}}(\boldsymbol{x}_i) = \frac{\boldsymbol{x}^{n+1}_i - \boldsymbol{x}^{n}_i}{\Delta t_{\text{dom}}}. \tag{8.51}$$

On the other hand, the acoustic model also requires an initial mesh (see Fig. 8.2) at the time instant $t = 0$. For 2D simulations this mesh corresponds to a rectangular domain with structured triangular elements of size $h \approx 0.001$ m, while for 3D simulations it is a cylinder with tetrahedra with $h \approx 0.002$ m (see Fig. 8.3). Their dimensions are obtained from the average values of the length and maximum radius of the five vowels vocal tract data set. However, the length has been finally fixed to L=18 cm (almost the mean length) for simplicity. These initial meshes for 2D and 3D can be viewed as being the vocal tract in a rest position (neutral vowel *schwa*).

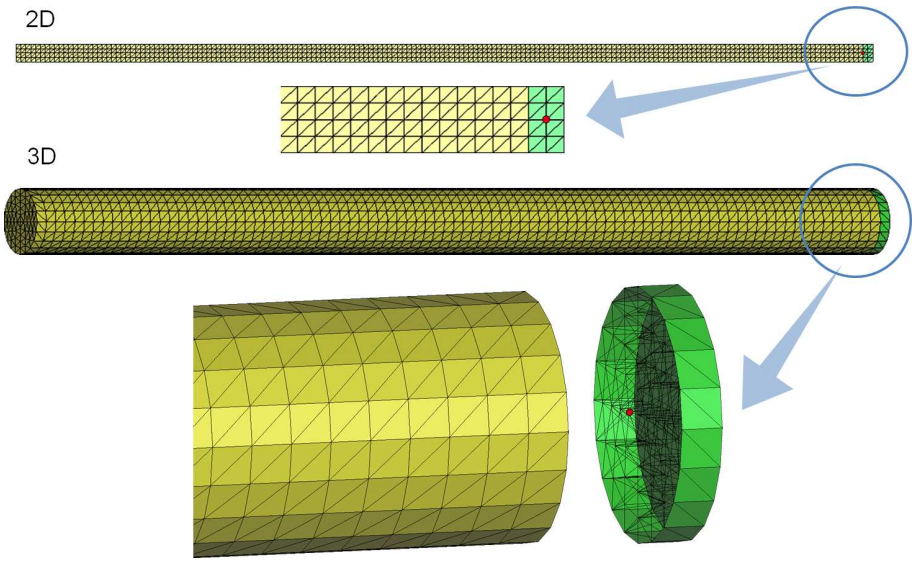
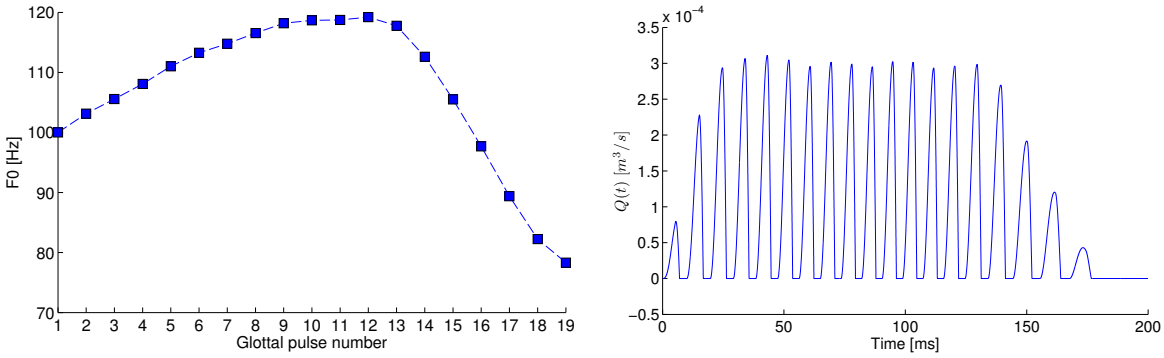


Fig. 8.3: Finite Element mesh for the 2D and 3D vocal tracts in the rest position. The red dot indicates the mesh node close to the mouth exit where the acoustic pressure is tracked for obtaining a diphthong sound.

In order to perform a diphthong sound, the corresponding initial mesh is gradually distorted to reach the first vowel target. During this process only the Laplacian equation (8.50) is computed to obtain the coordinates of the inner nodes. Once the mesh boundary has the shape of the first vowel, both the Laplacian equation (8.50) and the mixed wave equation in an ALE framework (8.5) are numerically solved. The 3D boundary node coordinates generated by the geometry model are linearly interpolated in time for simplicity. More elaborated interpolation models could be used instead to control the formant trajectory (see e.g. Story (2005), where a mapping between area functions and formant positions is used). However, the proposed approach to produce diphthongs will remain the same. The above strategy for the mesh movement has been found appropriate for diphthong production, since it shows little mesh distortion and thus avoids remeshing strategies, which are computationally expensive.

A speed of sound of $c_0 = 350$ m/s and an air density of $\rho_0 = 1.14$ kg/m^3 are considered in all simulations. For the time step, a value of $f_s = 1/\Delta t = 80$ kHz is used, which ensures enough time accuracy for the audible frequency range. The stabilization constants in Eq. (8.45) are taken as $C_1 = C_2 = 0.01$, which were found appropriate after some numerical tests. Concerning wall losses, a constant boundary admittance coefficient of $\mu = 0.005$ is considered, which is related to the wall impedance Z_w in Eq. (8.8b) by $\mu = \rho_0 c_0 / Z_w$.



(a) (b)

Fig. 8.4: (a) Fundamental frequency (F0) of each one of the glottal pulses generated by the 3D glottal model and (b) resulting volume velocity $Q(t) = S_g(t)g(t)$, with $S_g(t)$ standing for the glottal cross-sectional area and $g(t)$ for the particle velocity.

In what concerns the glottal source model, a train of glottal pulses is generated using a Rosenberg model of the C type (see Appendix B). In order to improve the naturalness of the generated sound, a pitch curve is followed to obtain the fundamental frequency of each glottal pulse (see Fig. 8.4a). In the particular case of a diphthong with a duration of 200 ms, this results in 19 glottal pulses that almost fill the duration of the diphthong sound. Moreover, some shimmer is also added to the amplitude of each pulse and a fade in/out is applied to the global signal. The resulting glottal source for this particular case can be observed in Fig. 8.4b. Note that there are not glottal pulses up to 200 ms in order to let the acoustic waves vanish within the vocal tract and therefore reach a zero pressure value.

The acoustic pressure is finally collected at a mesh node located 0.002 m from the vocal tract exit (see red dot in Fig. 8.3), and converted to an audio file to listen to the produced diphthong.

8.5 Numerical examples

8.5.1 Two-dimensional duct with periodically varying length

The proposed stabilized FEM formulation for diphthong production is first validated with the following numerical example. The test consists of a two-dimensional tube of constant cross-section that is being shortened and lengthened periodically by moving its boundary

exit (see Fig. 8.5). This example can be interpreted as a very rough approximation of a vocal tract in a rest position with a lip protrusion movement. At the beginning of the tube a sinusoidal velocity is imposed so that only a longitudinal resonance mode gets exited. Since for this mode plane waves will be generated within the tube, note that the mesh velocity $\boldsymbol{u}_{\text{dom}}$ in Eq. (8.5) will be parallel to both the gradient of the acoustic pressure and velocity, magnifying the importance of the ALE terms (i.e. in Eq. (8.5), the terms containing $\boldsymbol{u}_{\text{dom}} \cdot \nabla p$ and $\boldsymbol{u}_{\text{dom}} \cdot \nabla \boldsymbol{u}$).

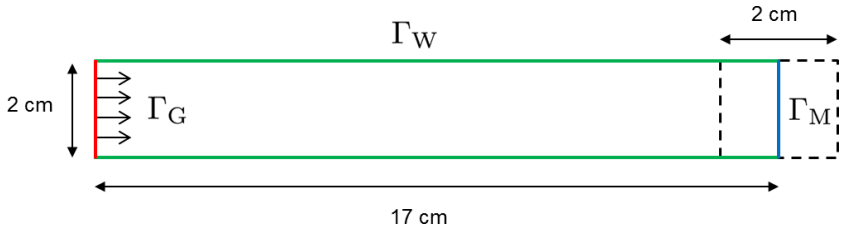


Fig. 8.5: Sketch of the tube example. Γ_G stands for the boundary where $g(t)$ is introduced, Γ_W for the tube walls and Γ_M for the mouth aperture, a contour that is moving following a sinusoidal pattern.

Let us split the boundary of the tube in three non intersection regions termed Γ_{G}, Γ_{W} and Γ_{M} (see Fig. 8.5). The boundary conditions that will be used to supplement the mixed wave equation in an ALE framework (8.5) will be

$$\boldsymbol{u} \cdot \boldsymbol{n} = g(t) \qquad \text{on } \Gamma_{\text{G}},\ t > 0, \tag{8.52a}$$

$$\boldsymbol{u} \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\text{W}},\ t > 0, \tag{8.52b}$$

$$p = 0 \qquad \text{on } \Gamma_{\text{M}},\ t > 0. \tag{8.52c}$$

The first boundary condition (8.52a) is used to impose a time varying velocity $g(t)$. Rigid walls are considered in Eq. (8.52b), while in Eq. (8.52c) an open-end boundary condition is imposed. As mentioned before, this tube is excited at a longitudinal resonance mode, in particular, the third resonance mode f_3 of the tube in a rest position, i.e. $g(t) = \sin(2\pi f_3 t)$. For a tube of length L with the boundary conditions described in Eq. (8.52) the frequency of the nth mode can be computed from (see e.g., Fletcher and Rossing, 1988)

$$f_n = (2n-1)\frac{c_0}{4L}. \tag{8.53}$$

So using $n = 3$, $c_0 = 350$ m/s and $L = 17$ cm we get that $f_3 \simeq 2573$ Hz. A numerical simulation has been performed with a sampling frequency of $f_s = 1/\Delta t = 80$ kHz. The mesh is updated at the same rate than the acoustic waves, i.e. $\Delta t_{\text{dom}} = \Delta t$.

The obtained results for the acoustic pressure at different time steps are shown in Fig. 8.6, considering the proposed stabilization (Fig. 8.6a) and with a non stabilized code (Fig. 8.6b). We can observe that in the first time instants, both solutions clearly form the typical pressure distribution of a 3th resonance mode. Two nodal lines and three regions of maximum/mimimum pressure can be appreciated. It should be noted that, although the tube length is changing in time, this variation is not large enough to produce a change of the resonance number and then modify the pressure pattern. The used input excitation has always a frequency close to the 3-th resonance, obtaining maximum pressure levels when the length is exactly L and minimum with $L = L \pm 0.01$ m. Comparing now both cases after some time steps (see in Fig. 8.6 t=23.45 ms), the numerical solution starts to become unstable if stabilization is not considered. If we let evolve the problem even more time (see in Fig. 8.6 t=25.8 ms, t=29.475 ms) the numerical solution clearly blows up without stabilization, totally exceeding the plotted range [-500, 500] Pa.

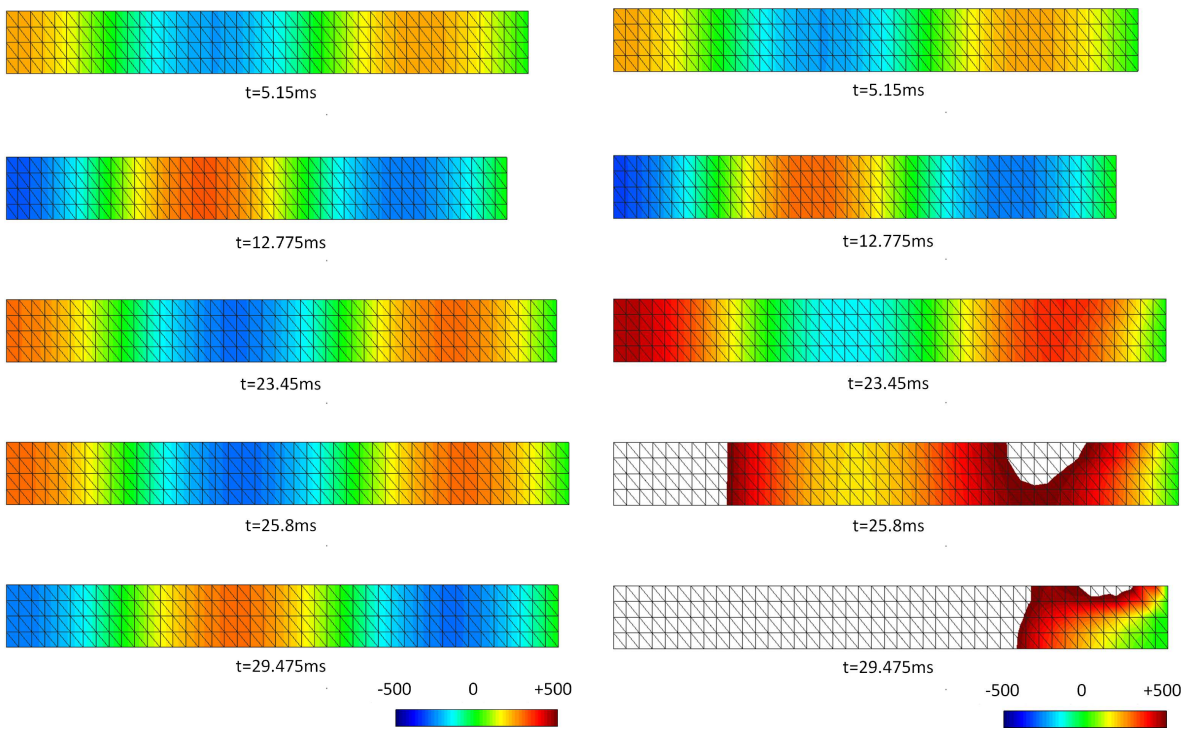


Fig. 8.6: Snapshots of the acoustic pressure in the tube example for different time instants considering the (a) stabilized and (b) the non stabilized formulation.

8.5.2 Production of diphthongs in 3D

A more applied example related to voice production will be next addressed. In particular, we will focus on the numerical generation of diphthongs. The diphthong /ai/ will be first generated for illustrative purposes though the presented procedure has general validity.

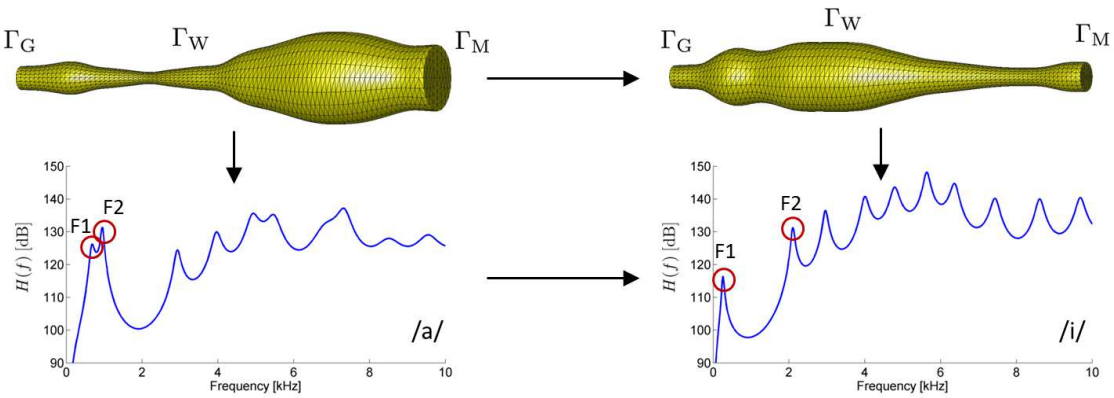


Fig. 8.7: Sketch for the numerical production of diphthong /ai/.

In the first part of the thesis, vowel sounds were produced by using simplified vocal tract geometries generated from the area functions in Story (2008) (see top row in Fig. 8.7 for vowels /a/ and /i/). After imposing a train of glottal pulses at the glottis area Γ_G, and impedance and radiation conditions at the vocal tract walls Γ_W and mouth exit Γ_M respectively, a vowel sound was produced by solving the irreducible wave equation (8.3) for the acoustic pressure (see Chapter 2). Additionally, one could also compute a vocal tract transfer function (see last row in Fig. 8.7) and obtain for instance the first two formants F1 and F2, which are the ones that allow listening individuals to distinguish between vowel sounds.

It is then clear that to generate the diphthong /ai/ we will have to transition from the situation in the left column in Fig. 8.7 to the one in the right column. At this point we would like to remark that some simplifications will be assumed. Since it is not the goal of this work to reproduce diphthongs from a given individual with high fidelity, we perform a linear interpolation from the geometry of the vocal tract of vowel /a/ to the vocal tract of vowel /i/, though more accurate options exist, based on controlling the formant transition trajectories between vowels Story (2005). Moreover, we simply impose a zero pressure release condition $p = 0$ at the mouth exit Γ_M. This is a rough approximation to emulate radiation losses, which depending on the formant-cavity affiliation phenomenon (Fant, 1970; Stevens, 2000), are known to shift the formants towards lower frequencies and to increase the formant bandwidths for mid-high frequencies (see e.g., Arnela *et al.*, 2013, and Chapter 5 and Chapter 6 for a detailed discussion).

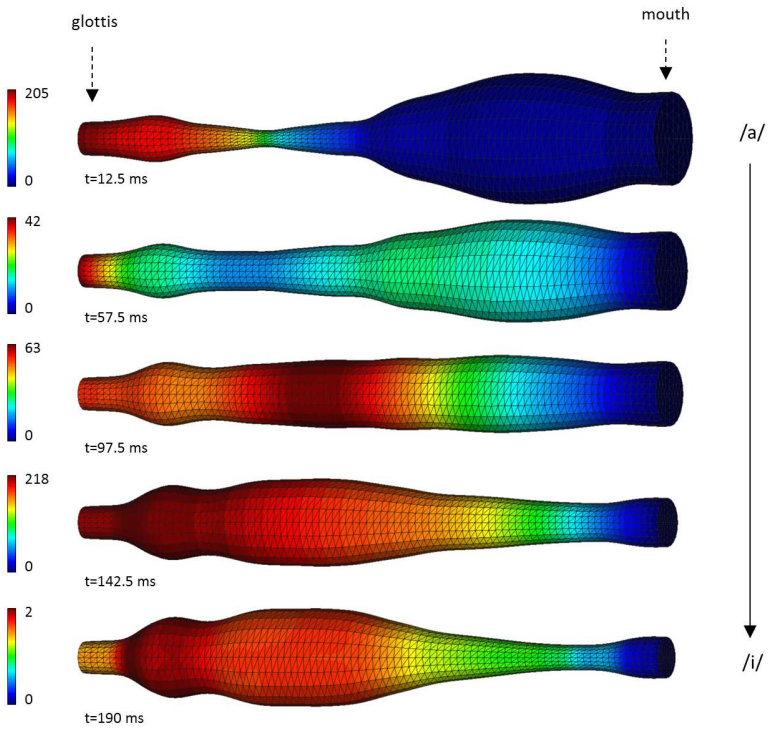


Fig. 8.8: Snapshots at five different time instants showing the transition from vowel /a/ to /i/. The acoustic pressure at the boundaries of the vocal tract and its shape evolution can be appreciated.

A more precise option would consist in allowing acoustic waves to emanate from the mouth and radiate towards infinity. This would involve, for example, implementing a perfectly matched layer (PML) (Berenger, 1994; Kaltenbacher *et al.*, 2013; Arnela and Guasch, 2013) for the finite element approximation of the ALE wave equation presented in this chapter, or imposing a good enough performing Sommerfeld boundary condition (see e.g., Espinoza *et al.*, 2014). This has been left for future developments.

To produce the diphthong /ai/ we solve the ALE mixed wave equation (8.5) supplemented with the boundary conditions in (8.8) using the stabilized FEM approach described in the Section 8.3, in the moving computational domain consisting of the transition from the vocal tract of /a/ in the left column of Fig. 8.7, to the geometry of /i/ in the right column. The moving vocal tract domain is initially meshed using a structured tetrahedral mesh with initial element size $h \simeq 0.002$ m and evolved according to the procedure described in Section 8.4. The mesh is updated in this case with $\Delta t_{\text{dom}} = 10\Delta t$, since the vocal tract moves very slowly compared to the speed of sound. The simulation last for 200 ms and a sampling rate of $f_s = 1/\Delta t = 80$ kHz is used.

In Fig. 8.8, a sequence of five snapshots corresponding to the transition from vowel /a/ to /i/ is presented. The changes in vocal tract shape can be clearly appreciated. The first snapshot at $t = 12.5$ ms corresponds to the articulation of an /a/, whereas the last one at $t = 190$ ms corresponds to that of an /i/. The acoustic pressure values at the vocal tract boundaries for the selected time instants can be also appreciated. Different limiting values are chosen for the color scales to better perceive the acoustic pressure distribution at the vocal tract for a particular time instant.

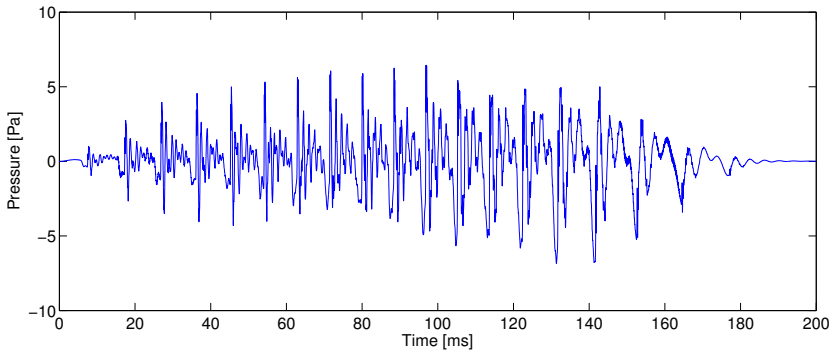


Fig. 8.9: Time evolution of the acoustic pressure at a point close to the mouth exit. When converted to audio, it produces the sound /ai/.

In Fig. 8.9 we show the time evolution for the acoustic pressure at a point close to the mouth exit. When transforming this plot into an audio file we get, as expected, the sound /ai/.

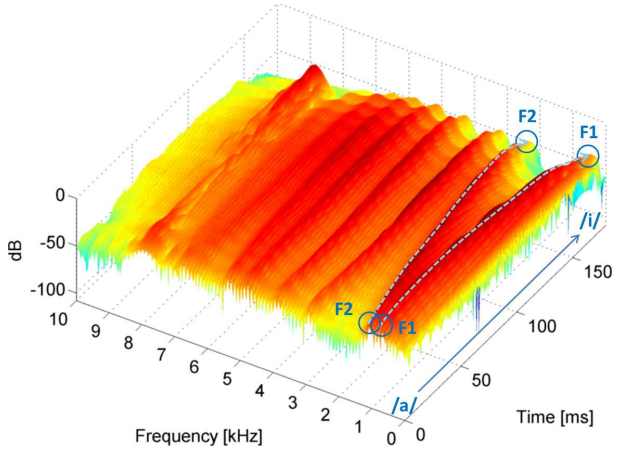
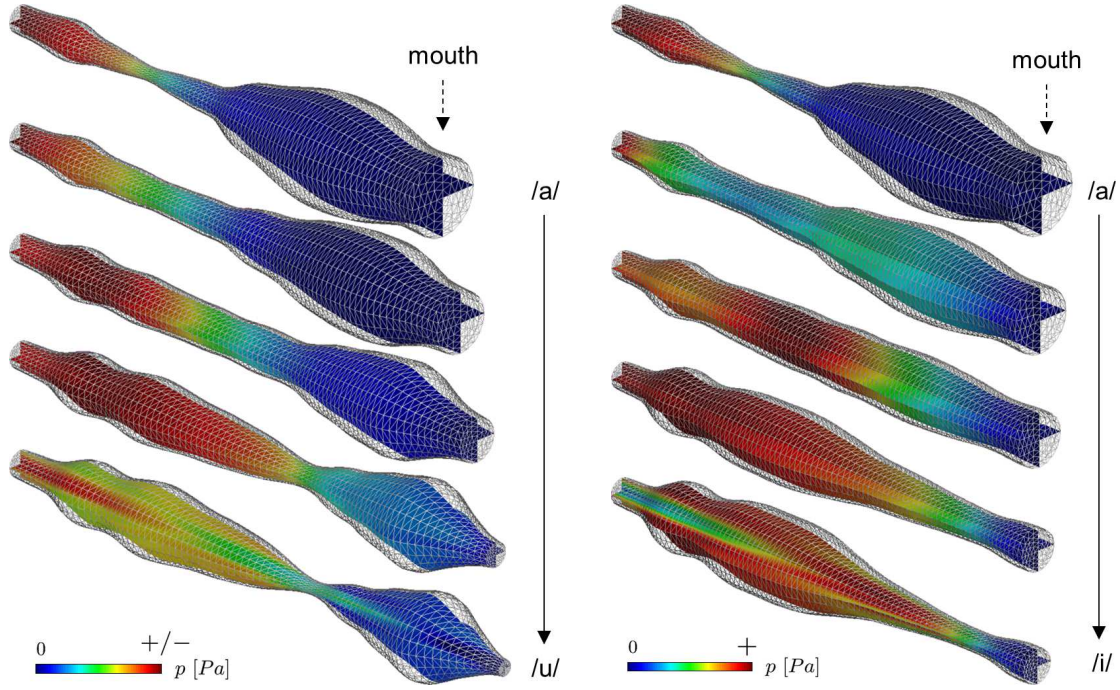


Fig. 8.10: Spectrogram of the acoustic pressure in Fig. 8.9 tracked at the node located close to mouth exit. Formant trajectories from /a/ to /i/ are highlighted.

Finally, in Fig. 8.10 we show the spectrogram corresponding to the acoustic pressure in Fig. 8.9. This is computed using a hamming window with a frame width of 20 ms and

an overlap of 1 ms. As usually done in speech processing, a pre-emphasis FIR filter with coefficients [1 -0.97] is applied to enhance the visualization of the high frequency range. In the first curve of the spectrogram, corresponding to vowel /a/, we can clearly identify the two formants previously highlighted in the vocal tract transfer function of /a/, in the last row of the left column of Fig. 8.7. Similarly, the last curve of the spectrogram corresponds to vowel /i/, and the two formants identified in the vocal tract transfer function of /i/ (last row of the right column of Fig. 8.7) can also be clearly recognized. The trajectories of the two formants F1 and F2, which separate apart from /a/ to /i/, are indicated in Fig. 8.10 to better show their evolution.

A second example is carried out to generate the diphthong /au/. Some snapshots are presented in Fig. 8.11 showing the transition from vowel /a/ to vowel /u/ (Fig. 8.11a). In

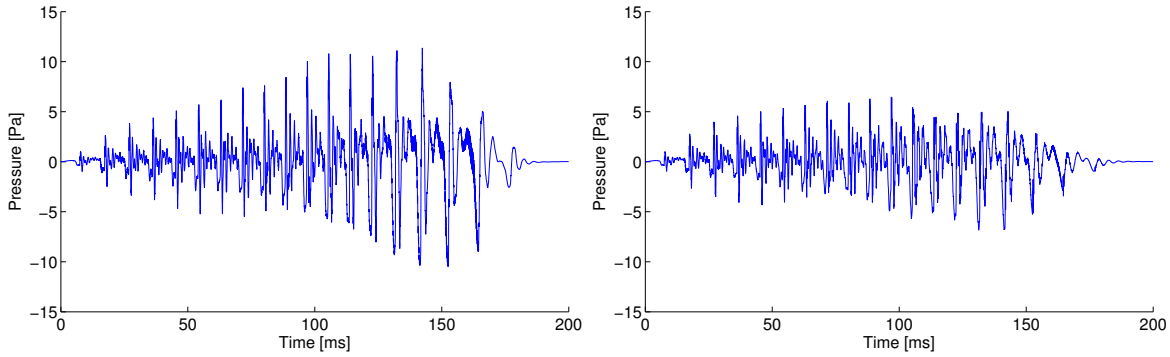


(a) /au/ (b) /ai/

Fig. 8.11: Snapshots showing the acoustic pressure p within two cuts of the 3D vocal tract for diphthongs /au/ and /ai/ at $t = 12.5,\ 57.5,\ 97.5,\ 142.5$ and 190 ms; the first time instant corresponds to the first vowel, the last one to the second vowel, and the intermediate ones to the transition between vowels. In each frame the color scale is tuned to the maximum absolute value ("+/−" in the legend) within the vocal tract, which corresponds to an acoustic pressure value of 202, 144, 120, 188 and −3 Pa for diphthong /au/, and 205, 41, 63, 218 and 2 Pa for diphthong /ai/, at time instants t=12.5, 57.5, 97.5, 142.5 and 190 ms respectively.

addition, diphthong /ai/ is also included in the figure for comparison purposes (Fig. 8.11b). The vocal tract geometry is represented by the outer mesh while the acoustic pressure is plotted in two cuts in the midsagittal and midtransverse planes, which allow for a better visualization of the acoustic field in the interior. Positive and negative values for the acoustic pressure can be produced within the vocal tract. In the case of diphthong /ai/ all the presented snapshots have positive values (denoted by "+" in the legend). However, for diphthong /au/ some frames can contain either positive or negative values (denoted by "+/−" in the legend). In this case only the last frame corresponding to vowel /u/ has negative values in the interior. Note however that the zero pressure value is always denoted with the same color and therefore, that the acoustic pressure at the mouth has always this color given that at this cross-section an open-end boundary condition is prescribed, see Eq. (8.8c). On the other hand, the maximum absolute value of the color scale is adapted to each frame to improve the visualization of the acoustic waves inside the vocal tract. Although at first sight one could think that only plane waves are produced within the vocal tract, we can observe in the last snapshot corresponding to vowels /i/ and /u/ that a complex field is generated. However, it has to be noted that at this time instant the glottal model is not active (see Fig. 8.4b at $t = 190.2$ ms) so the acoustic pressure within the vocal tract is really small (maximum/minimum of 2/-3.4 Pa for /ai/ and /au/ respectively), compared to other time instants (maximum values ranging from 41 to 218 Pa).

In order to listen to the generated diphthong, the acoustic pressure is tracked at a node close to the mouth exit (see Fig. 8.3). The obtained signals for diphthongs /au/ and /ai/ can be observed and compared in Fig. 8.12. One can appreciate in the figure how the time signal smoothly transition from one vowel to the other.



(a) /au/ (b) /ai/

Fig. 8.12: Acoustic pressure evolution captured at a node close to the vocal tract mouth (see red dot in Fig. 8.3) for the 3D diphthongs /au/ and /ai/.

8.5.3 Production of diphthongs in 2D

Finally, diphthong sounds are generated following a two-dimensional (2D) approach. To do so, the 2D vocal tract methodology for static vowel sounds presented in Chapter 7 is extended to the production of diphthongs. Following a four step methodology, we showed that the acoustic behavior of three-dimensional vocal tracts with circular cross-sections could be recovered to some extent by appropriate modifications of two-dimensional vocal tracts. The first three steps in Arnela and Guasch (2014b) (see Chapter 7) will then be adapted to generate diphthong sounds. Following this methodology 3D formant positions, bandwidths and energy can be respectively recovered by means of a tuning process of the 2D vocal tract geometry, the 2D boundary admittance coefficient and the 2D glottal flow. Otherwise, if the standard 2D approach is followed (based on performing a midsagittal cut of a 3D vocal tract with circular cross-sections, see e.g., Mullen *et al.*, 2006, 2007; Speed *et al.*, 2009), significative errors are produced in the vowel formants (Arnela and Guasch, 2014b,a). With regards to the last fourth step concerning the correction of radiation losses, it will not be considered herein for simplicity. Free-field radiation will be included in future works.

The methodology, as explained in Chapter 7, consists in tuning the 2D vocal tract radius function $a_{\mathtt{2D}}(x,t)$, the boundary admittance coefficient $\mu_{\mathtt{2D}}(x,t)$ and the glottal acoustic velocity $g_{\mathtt{2D}}(t)$ (see Fig. 8.13). Our goal is to reproduce the acoustics of a dynamic 3D vocal tract with circular cross-sections built from the interpolation of the area functions $S_{\mathtt{3D}}(x,t) = \pi a_{\mathtt{3D}}^2(x,t)$ in Story (2008), with a constant boundary admittance coefficient of $\mu_{\mathtt{3D}} = 0.005$ imposed at the walls, and with a particle velocity $g_{\mathtt{3D}}(t)$ generated by the vocal folds introduced at the glottal cross-section $S_{g,\mathtt{3D}}$.

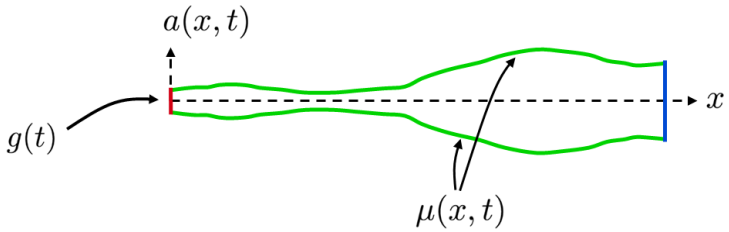


Fig. 8.13: Dynamic 2D vocal tract model, with $a(x,t)$ standing for the radius of the vocal tract cross-sections, $\mu(x,t)$ for the boundary admittance of the walls and $g(t)$ for the acoustic velocity generated by the glottal folds.

In the first step, the 2D radius function $a_{\mathtt{2D}}(x,t)$ is modified so that 3D formant positions are recovered. To do so, at a time instant n the following recursive equation is used

$$a_{\mathtt{2D}}(x_i^n) = \begin{cases} a_{\mathtt{2D}}(x_{i+1}^n)\ m(x_i^n) & 1 \leq i < d \\ a_{\mathtt{2D}}(x_d^n) & i = d \\ a_{\mathtt{2D}}(x_{i-1}^n)\ m(x_i^n) & d < i \leq N \end{cases}, \tag{8.54a}$$

with

$$m(x_i^n) = \begin{cases} S_{\mathtt{3D}}(x_i^n)/S_{\mathtt{3D}}(x_{i+1}^n) & i < d \\ S_{\mathtt{3D}}(x_i^n)/S_{\mathtt{3D}}(x_{i-1}^n) & i > d \end{cases}. \tag{8.54b}$$

In Eq. (8.54) x_d^n is a point inside the vocal tract $x_1^n < x_d^n < x_N^n$, with x_1^n and x_N^n standing for points at the glottal and mouth cross-sections at time instant n, respectively. The point x_d^n is used to provide an initial condition to the recursive equation (8.54), and corresponds to the distance from where the maximum vocal tract cross-sectional area (maximum $a_{\mathtt{3D}}^n$) is produced with respect to the glottis. After imposing that the onset of higher order modes at this point has to be the same in 2D and 3D, we get that $a_{\mathtt{2D}}(x_d^n)$ can be computed as

$$a_{\mathtt{2D}}(x_d^n) = a_{\mathtt{3D}}(x_d^n)\frac{0.5\pi}{1.84}. \tag{8.55}$$

The second step of the approach is based on tuning the wall boundary admittance $\mu_{\mathtt{2D}}$ to mimic the 3D formant bandwidths. This can be done by equating the 2D and 3D axial wave numbers at each time instant, which yields

$$\mu_{\mathtt{2D}}(x_i^n) = \mu_{\mathtt{3D}}(x_i^n)\frac{2a_{\mathtt{2D}}(x_i^n)}{a_{\mathtt{3D}}(x_i^n)}, \quad 1 < i \leq N. \tag{8.56}$$

In the third step, we impose that the particle velocity $g(t)$ at the glottal cross-section to be the same in 3D and 2D. This leads to

$$g_{\mathtt{2D}}(t^n) = \frac{Q_{\mathtt{3D}}(t^n)}{S_{g,\mathtt{3D}}(t^n)}, \tag{8.57}$$

where $S_{g,\mathtt{3D}}$ is the 3D glottal cross-sectional area and $Q_{\mathtt{3D}}$ is the 3D volume velocity typically generated by the glottal models (e.g., Rosenberg, 1971).

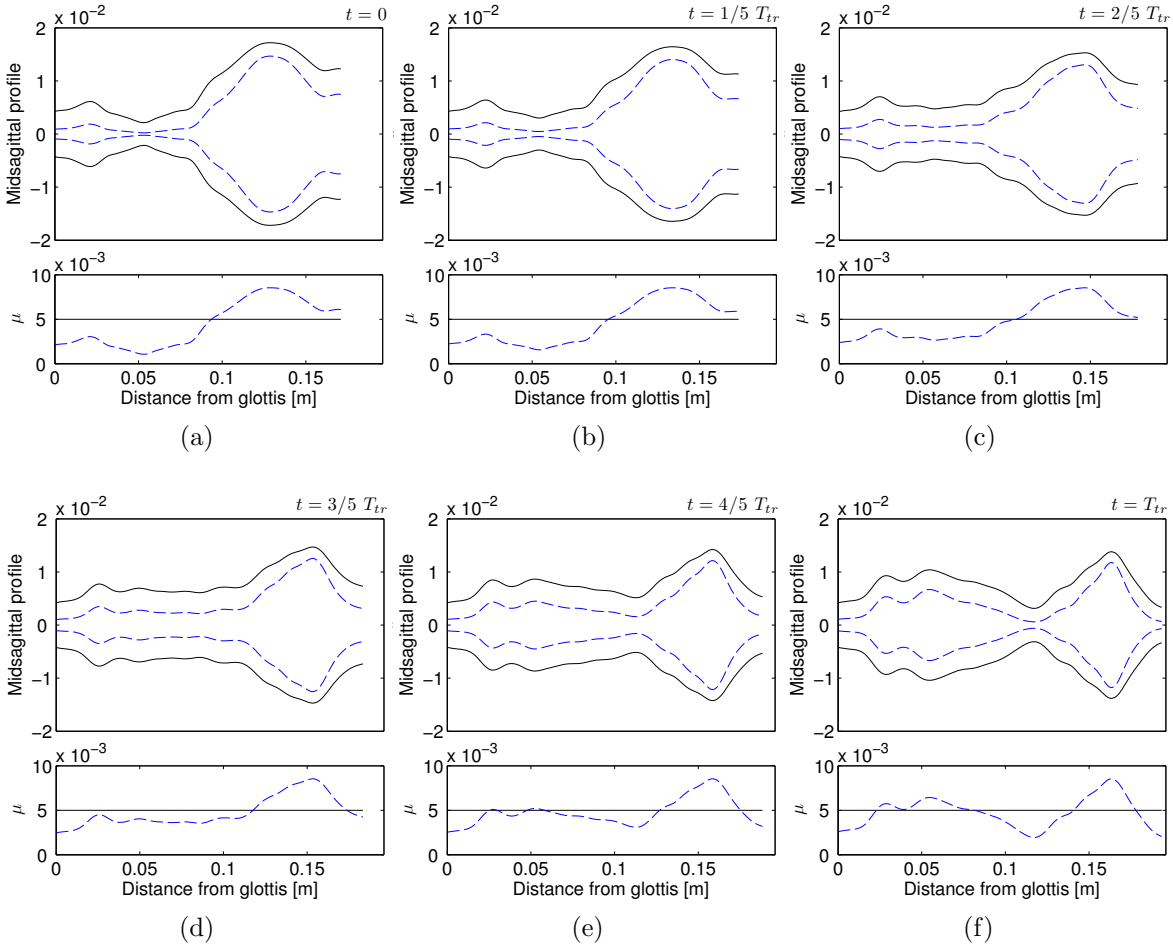


Fig. 8.14: Midsagittal profiles and boundary admittance coefficients μ of diphthong /au/ at time instants $t = i/5\ T_{tr}$, where T_{tr} is the transition time and $i = 0, 1, 2, 3, 4, 5$. Linear interpolation is used in the 3D vocal tracts. (— 3D, - - 2D).

In order to generate a diphthong sound, the three-dimensional radius function a_{3D} is first linearly interpolated, for simplicity, to produce the transition from one vowel to the other. The area functions in Story (2008) are used for the static vowels. Then, a_{2D} is obtained by applying Eq. (8.55), μ_{2D} is computed with Eq. (8.56) and g_{2D} with Eq. (8.57). In Fig. 8.14 an example for diphthong /au/ is shown for the six time instants $t = i/5\ T_{tr}$, with T_{tr} being the transition time and $i = 0, 1, 2, 3, 4, 5$. In what concerns wall losses, note from Eq. (8.56) that, although they are time independent in 3D, this is not the case in 2D because μ_{2D} evolves in time given that $a_{\text{2D}}/a_{\text{3D}} \neq 1$ (see Fig. 8.14).

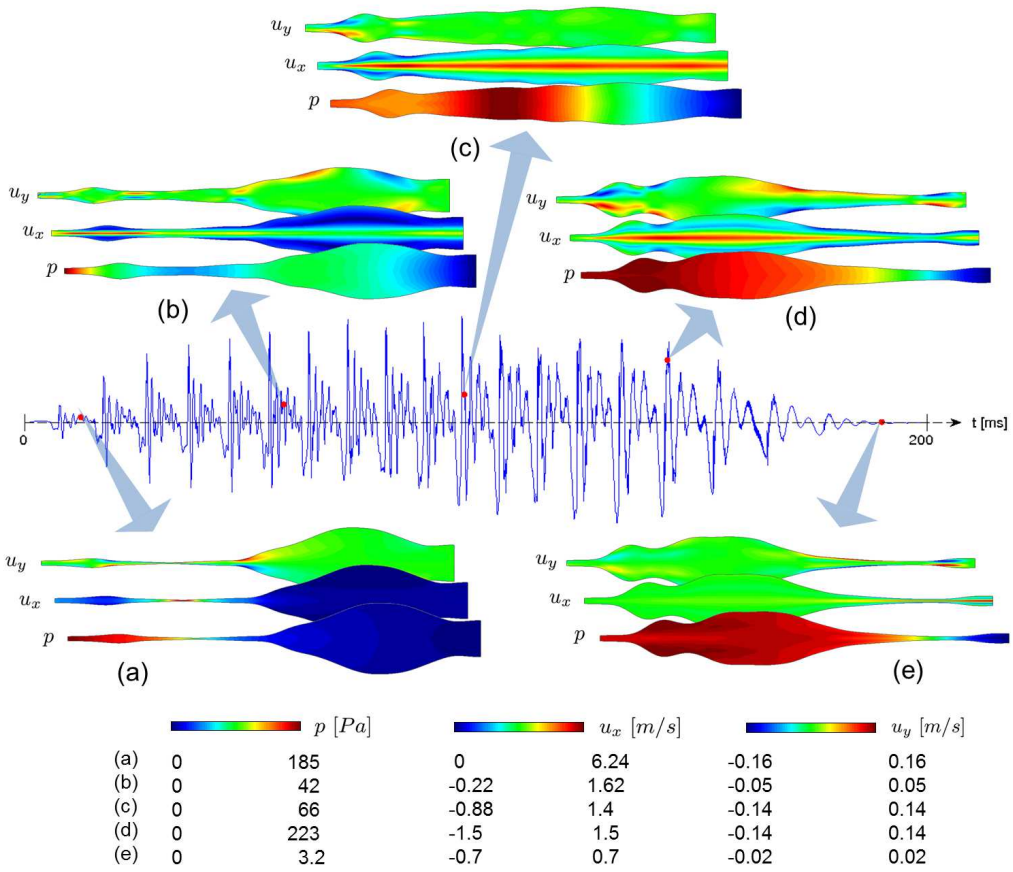
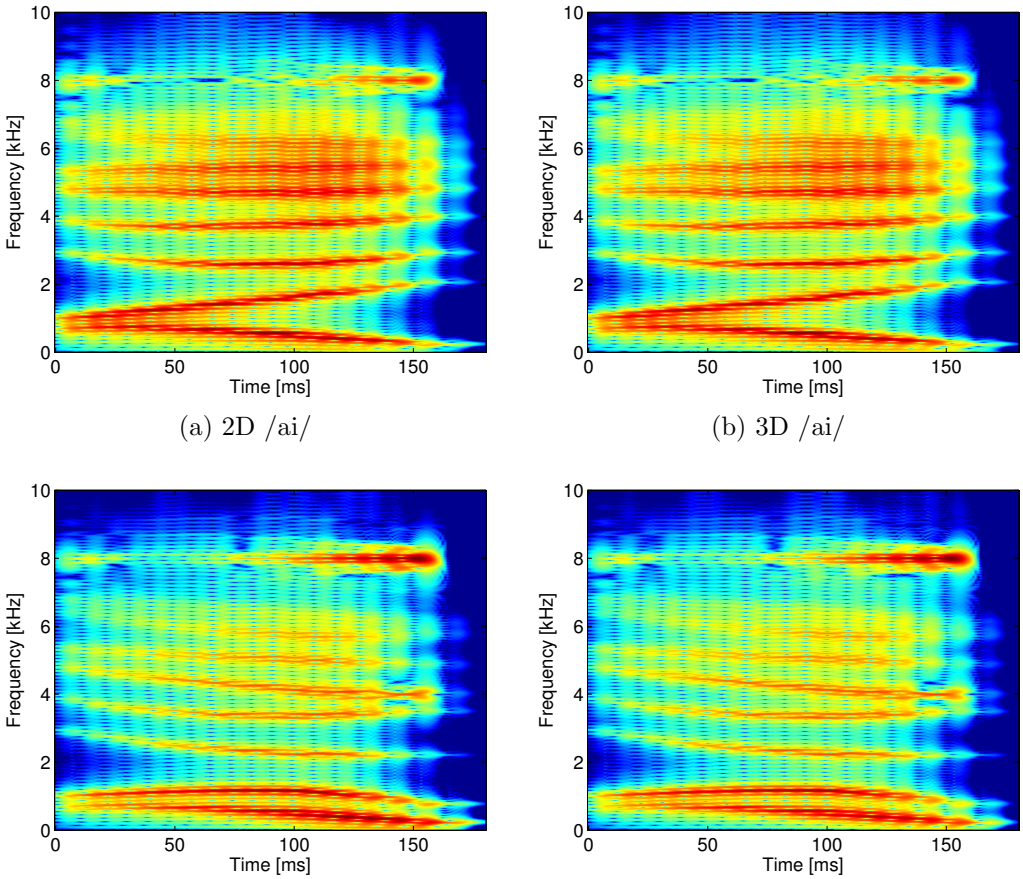


Fig. 8.15: Acoustic pressure evolution of the 2D diphthong /ai/ collected at a node close to the mouth exit (see Fig. 8.3), with some snapshots showing the acoustic pressure p and particle velocity components, u_x and u_y, within the vocal tract. Values are taken at (a) $t = 12.5$ ms (vowel /a/), (b)(c)(d) $t = 57.5,\ 97.5,\ 142.5$ ms (transition from /a/ to /i/) and (e) $t = 190$ ms (vowel /i/). The color scale is adapted in each frame.

In Fig. 8.15 some snapshots are presented for the 2D diphthong /ai/. The distribution of the acoustic pressure p and the particle velocity components v_x and v_y are shown at the same five time instants than in the 3D case (see Fig. 8.11). The first one corresponds to an instant during the articulation of vowel /a/ ($t = 12.5$ ms), the second, third and forth to an instant during the transition from /a/ to /i/ ($t = 57.5,\ 97.5,\ 142.5$ ms), and the fifth one to an instant once we get vowel /i/ ($t = 190$ ms). The acoustic pressure evolution of the generated diphthong is also represented to better illustrate the acquisition times of these snapshots. Complex patterns for the pressure and velocity fields can be appreciated within the 2D vocal tract, which could not certainly be attained from 1D models because they assume plane wave propagation. Similar pressure distributions are obtained compared with the 3D snapshots in Fig. 8.11.



(a) 2D /ai/ (b) 3D /ai/

(c) 2D /au/ (d) 3D /au/

Fig. 8.16: Spectrogram of the simulated diphthongs /ai/ (top) and /au/ (bottom) for 2D (left) and 3D (right).

In Fig. 8.16 the spectrogram of the generated diphthongs can be observed for both 3D and 2D. These are computed from the acoustic pressure signal captured at a node close to the mouth cross-section (red dot in Fig. 8.3). As usually done in speech analysis, a pre-emphasis filter is applied to enhance the visualization of higher frequencies during the spectrogram computation. One can observe in the figure that the formants characterizing vowel /a/ smoothly transition to those of vowel /i/ for /ai/ and to those of vowel /u/ for /au/. Comparing the 3D and 2D spectrograms a very similar formant trajectory can also be appreciated. Informal perceptual tests in fact have shown that no significative difference is produced between the 3D and the 2D diphthong sounds. Therefore, the voice quality is preserved in 2D, but with the advantage of achieving a large reduction of the computational time compared to 3D. The performed 2D simulations have last about ~ 8 minutes while ~ 8 hours have been needed for 3D computations in a regular desktop computer (Intel(R) Core(TM)i5 2.8 GHz).

8.6 Conclusions

In this chapter a FEM approach for the numerical production of diphthongs has been proposed. This has consisted in solving the mixed wave equation in an ALE frame of reference, rather than working with its irreducible counterpart, usually employed in vowel production. However, the numerical resolution of the former is more intricate than that of the latter and requires stabilization in order to use equal interpolation for the acoustic pressure and velocity fields. An algebraic subgrid scale strategy has been followed for this purpose. The motion of the Finite Element meshes has been addressed by solving with FEM the Laplacian equation, which smoothly translate the prescribed movement of the wall nodes to the inner nodes through diffusion. A strategy based on building an initial rectangular mesh emulating the vocal tract in a rest position has been also presented as a useful methodology that avoids remeshing strategies an therefore minimizes the computational cost. The proposed stabilized FEM formulation has been first checked with a numerical benchmark problem consisting of plane waves propagating in a two-dimensional duct with a moving exit boundary. Then, the diphthongs /ai/ and /au/ have been generated as examples using a 3D and a 2D approach. For the latter, a dynamic 2D vocal tract model has been constructed following three simple steps. In the first step the vocal tract geometry has been modified to adjust 3D formant positions, in the second one the boundary admittance coefficient has been tuned to get 3D formant bandwidth and energy, and in the third one, the glottal velocity has been adapted to recover 3D pressure levels. A very similar smooth formant transition to that in 3D has been reproduced in 2D. Moreover, the dynamic 2D model has shown complex inner vocal tract acoustic pressure and velocity patterns that are also produced in 3D and that cannot be obviously given in 1D. Audio files have been generated presenting no significant differences between 3D and 2D from a perceptual point of view, and showing fairly good quality, even though simple homogeneous Dirichlet conditions have been used for the acoustic pressure at the mouth exit. Future work will extend the presented approach to include outward sound wave propagation so as to account for radiation losses. Moreover, more complex geometries generated e.g. from MRI will be considered.

Part IV

Concluding remarks

Chapter 9

Conclusions and future work

9.1 Conclusions

This thesis has focused on the modelling of the vocal tract acoustics by means of numerical methods, in particular the Finite Element Method. Conclusions at the end of each chapter have been already provided, so in this section the main contributions will be only outlined.

The first part of the thesis has dealt with the numerical production of vowels. The main contributions can be summarized as follows:

1. **A Finite Element based strategy for vowel production** has been presented in Chapter 2. This has consisted in solving the time domain wave equation for the acoustic pressure supplemented with appropriate boundary conditions for voice generation. A Perfectly Matched Layer has been used to absorb the sound waves emanated from the mouth, allowing us to consider free-field radiation conditions. The proposed approach has been validated against experiments in Chapter 3.

   The FEM formulation has been published in (J1) while the validation against experiments has resulted in (J4), (C4), (C6) and (C7), see Section 9.3.

2. **A strategy for the numerical computation of vocal tract impedances** based on an experimental transfer function method (TMTF) has been proposed in Chapter 4. This methodology allows us to compute an impedance with the sole use of the acoustic pressure (the acoustic particle velocity is not required), which makes it suitable for the proposed FEM formulation in Chapter 2.

   This work has been presented in the publications (J1) and (C1), see Section 9.3.

3. **A study of the acoustic effects of human head geometry simplifications in numerical simulations** has been performed. Once analyzed in Chapter 5 the radiation effects of vowel production, in Chapter 6 it has been shown that a spherical head with lips behaves similar to a realistic head up to 10 kHz from a perceptual point of view, which can save a lot of time during the geometry modelling. On the other hand, it has been observed that, for the considered head geometry, the lips can be removed from the spherical baffle if one limits the analysis up to 5 kHz. As a consequence, the theoretical load model of the piston set in a spherical baffle, widely used to speed up simulations, has been validated for the $[0, 5]$ kHz frequency range.

   This has yield the publications (J2), (C2) and (C5), see Section 9.3.

4. **A two-dimensional numerical approach with three-dimensional acoustic behavior** has been developed in Chapter 7 to synthesize vowel sounds. This has consisted of a four step methodology in which the parameters of a two-dimensional vocal tract model have been tuned to recover to a large extent the 3D formant positions, formant bandwidths, formant energies and pressure levels of a 3D vocal tract with circular cross-sections. Otherwise, if the standard 2D approach is followed, which considers a midsagittal cut of the above 3D vocal tract to generate its 2D counterpart, strong errors are produced in the vowel formants. The presented methodology solves this problem and results in a very good balance between the computational cost and the generated voice quality.

   This has been reflected in the publications (J3) and (C9), see Section 9.3.

In the second part of the thesis the numerical generation of diphthongs has been addressed. The main contributions made in this part are:

5. **A stabilized Finite Element formulation for diphthong production** has been derived in Chapter 8 to account for moving vocal tracts. The mixed wave equation for the acoustic pressure and acoustic particle velocity has been used in this occasion, which has been found more appropriate than the irreducible wave equation once expressed in an ALE frame of reference. However, this equation suffers from stability issues when the same interpolation is used for the acoustic pressure and particle velocity. For that purpose, a stabilized Finite Element approach based on an algebraic subgrid scale strategy has been developed. Finally, some diphthong sounds have been generated with simplified three-dimensional geometries. Moreover, the 2D approach presented in Chapter 7 has been extended to consider dynamic vocal tracts. Two-dimensional diphthong sounds have been produced, with almost the same voice quality than 3D ones.

   This has resulted in the publications (J5), (C8) and (C10), see Section 9.3.

9.2 Future work

The work performed in this thesis leaves some future lines for exploration, being some of them under current development.

With regard to the generation of vowels,

- **Simplifications of the vocal tract geometry** such as those considered in Section 2.5.1 may be studied in further detail. On the one hand, this would contribute to a better understanding of the complex 3D acoustic field within the vocal tract. On the other hand, these simplified geometries might facilitate voice simulation in a wide range of applications (e.g., interpolation of static vocal tract geometries to generate a sound transition, 3D multimodal approaches which need analytically defined contours, etc.). This subject is currently being performed.

- **Frequency dependent wall impedances** should be implemented for time domain simulations to replace the current constant value ones. This would not only enhance the modelling of the vocal tract wall losses but also could allow us to simply impose load impedance theoretical models on the mouth aperture to emulate radiation losses, instead of considering the more computational expensive, but also more accurate free-field radiation solutions. To do so, a formal convolution integral induced by the inverse Fourier transform of the impedance condition should be performed along the related contour, requiring large computational times and the storage of the unknowns along many time steps. This could be circumvented to some extent following for instance the strategy in Nieuwenhof and Coyette (2001), where the frequency dependent impedance is approximated in a rational form which yields a more efficient discrete representation of the convolution.

- **Calculation of high order modes direct radiation impedances** for several configurations of the human head geometry is also an interesting research line that requires further exploration. These are needed to correctly characterize the radiation impedance above the plane wave cutoff frequency, and could be used for instance as load impedances in three-dimensional approaches. Some theoretical models can be found in literature (e.g., Zorumski, 1973; Kemp *et al.*, 2001). However, they relay on many approximations, such as using an infinite flat baffle for the human head shape or removing the lips from the model, though arbitrary shaped exits can be achieved numerically (see e.g., Blandin *et al.*, 2014a). It would be then interesting to extend the current work so as to see how modal radiation impedances get modified with human head geometry simplifications.

- **The two-dimensional approach could be improved** by considering bent vocal tracts, side branches such as the piriform fossae and valleculae or asymmetrical boundaries. Each of the above configurations will suppose a new challenging problem, since these elements produce the onset of further resonances and antiresonances which may correspond to the three-dimensional ones.

- **The acoustic interaction between the vocal folds and the vocal tract** also constitutes a challenging and interesting research line to explore. This could be considered by adapting for instance a two-mass model to the 3D computational approach, or by directly simulating in the same computational domain the fluid-structure interaction of the vocal folds and the acoustics of the vocal tract. However, the latter requires solving more complex equations such as the compressible Navier-Stokes equations, consider fluid-structure strategies, and make use of acoustic analogies.

Concerning the generation of diphthongs,

- **Non reflecting boundary conditions** still need to be considered for this particular case to account for acoustic radiation of sound waves emanating from the mouth. The Sommerfeld boundary condition in Espinoza *et al.* (2014) is currently being implemented for that purpose.

- **Flexible boundaries** may be taken into account to enhance the wall loss mechanism. At present, only rigid walls are being considered. The walls elastic back reaction due to the inner acoustic pressure fluctuations will allow to better emulate the behavior of a realistic vocal tract (see e.g., Doel and Ascher, 2008, where this is done for 1D). This topic is also being investigated.

- **The use of realistic dynamic vocal tract geometries** to generate diphthong sounds presents a challenging problem to be solved. On the one hand, to avoid strongly distorted elements, remeshing strategies may be required to deal with dynamic realistic vocal tract. On the other hand, one has to generate these moving geometries. Since obtaining them from real-time MRI is still a subject for further exploration, an alternative may consist in resorting to simplified vocal tract geometries which can be easily interpolated to generate a transition. This has been the option adopted in this work, but it could be improved by using enhanced simplified vocal tracts e.g., incorporating bending and elliptical cross-sections. This strategy is under study.

Although not addressed in this work, the next natural step in computational voice production may be that of dealing with the production of syllables. New geometrical issues and physical phenomena will have to be considered. For example, in order to generate a syllable containing a nasal consonant (e.g., /na/), the nasal cavity geometry will have to be joined to the vocal tract. Stronger difficulties will be encountered in the simulation of syllables containing fricatives (e.g., /sa/), because the aerodynamic sound generated by the airflow vorticity leaving the lips at the output of the vocal tract, will have to be computed. Consequently, it will be necessary to resort to computational aeroacoustics.

As a further remark, the research done in computational voice production should contribute to build a full 3D voice simulation engine, which may enlarge the current knowledge on voice generation and help to better understand its production mechanisms. However, it seems clear that this will need huge computational resources and that supercomputer facilities will be probably required. In parallel to this development, one should explore alternatives to speed up these simulations to make the voice simulation engine easier to handle by the potential users. Simplifications on the glottal source models, vocal tract geometry and vocal tract acoustics models could be a possible strategy to reach this goal. Needless to say that the performance of these simplifications should be checked, when possible, against the full 3D version of the model. Of course, every time that a simplification is performed, some voice quality is lost, but the challenge here will be that of finding the configuration that provides the best balance between the computational cost and the voice quality (in objective or subjective terms), which may differ on the specific application and user demands.

Finally, it is worthwhile mentioning that other research fields of acoustics share similar problems to those of voice production. This is the case of musical acoustics of wind instruments (see e.g., Fabre *et al.*, 2012). For instance, in a brass instrument (e.g., a trumpet), the vibration of the player's lips can be interpreted as the vocal fold oscillation and the propagation of the generated sound waves trough the pipe, which latter radiate to free-field from the bell (exit), behaves similar to the vocal tract acoustics. In the case of woodwind instruments (flutes and reed instruments), the pipe acoustics could be emulated in a similar way to the vocal tract, but using a different aeroacoustics excitation mechanism. Although less related, other fields such as room acoustics could take benefit from the techniques developed in vocal tract acoustics simulation, since they also have to deal with complex geometries (i.e. rooms with a complex shape) and might be interested in time domain solutions of acoustic wave propagation (e.g., in auralization).

9.3 Publications, grants and awards

Academic journals

(J5) Oriol Guasch, Marc Arnela, Ramon Codina, Hector Espinoza, "A stabilized finite element method for the mixed wave equation in an ALE framework with application to diphthong production," Journal of Computational Physics, Submitted.

(J4) Rémi Blandin, Marc Arnela, Oriol Guasch, Rafael Laboissière, Xavier Pelorson, Annemie Van Hirtum and Xavier Laval, "Effects of higher order propagation modes in vocal tract like geometries," Journal of the Acoustical Society of America, Accepted.

(J3) Marc Arnela and Oriol Guasch (2014), "Two-dimensional vocal tracts with three-dimensional behavior in the numerical generation of vowels," Journal of the Acoustical Society of America, 135(1), pp. 369-379.

(J2) Marc Arnela, Oriol Guasch and Francesc Alías (2013), "Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations," Journal of the Acoustical Society of America, 134(4), pp. 2946-2954.

(J1) Marc Arnela and Oriol Guasch (2013), "Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method," Journal of the Acoustical Society of America, 133(6), pp. 4197-4209.

International conferences

(C10) Marc Arnela, Oriol Guasch, Ramon Codina and Hector Espinoza (2014), "Finite element computation of diphthong sounds using tuned two-dimensional vocal tracts," 7th Forum Acusticum, September 7–12, Krakow, Poland.[1]

(C9) Marc Arnela and Oriol Guasch (2014), "Three-dimensional behavior in the numerical generation of vowels using tuned two-dimensional vocal tracts," 7th Forum Acusticum, September 7–12, Krakow, Poland.[1]

[1]EAA Best Paper and Presentation Award for young researchers.

(C8) Oriol Guasch, Ramon Codina, Marc Arnela and Hector Espinoza (2014), "A stabilized Arbitrary Lagrangian Eulerian finite element method for the mixed wave equation with application to diphthong production," 11th Word Congress on Computational Mechanics (WCCM XI), July 20–25, Barcelona, Catalonia, Spain.

(C7) Rémi Blandin, Xavier Pelorson, Annemie Van Hirtum, Rafael Laboissière, Marc Arnela and Oriol Guasch (2014), "Modélisation aéroacoustique de la production de la parole," Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal audio, July 2–4, Lyon, France.

(C6) Rémi Blandin, Xavier Pelorson, Annemie Van Hirtum, Rafael Laboissière, Oriol Guasch and Marc Arnela (2014), "Effet des modes de propagation non plan dans les guides d'ondes à section variable," French Acoustical Conference 2014, April 22–25, Poitiers, France, pp. 745–751.

(C5) Marc Arnela and Oriol Guasch (2014), "Validation of the piston set in a sphere model for vowel sound radiation losses against realistic head geometry using time-domain finite-element simulations," 9th International Conference on Voice Physiology and Biomechanics (ICVPB), April 10–12, Salt Lake City, USA.

(C4) Xavier Pelorson, Annemie Van Hirtum, Boris Mondet, Oriol Guasch and Marc Arnela (2013), "Three-dimensional vocal tract acoustics," Acoustics 2013, November 10–15, New Delhi, India.

(C3) Oriol Guasch, Sten Ternström, Marc Arnela and Francesc Alías (2013), "Unified numerical simulation of the physics of voice. The EUNISON project," ISCA 8th Speech Synthesis Workshop (SSW8), Demo Session pp. 241-242, August 31–September 2, Barcelona, Catalonia, Spain.

(C2) Marc Arnela, Oriol Guasch and Francesc Alías (2012), "Analysis of the radiation effects on vowels by means of time domain finite element simulations," 19th International Congress on Sound and Vibration (ICSV19), July 8–12, Vilnius, Lithuania.

(C1) Marc Arnela and Oriol Guasch (2012), "Adaptation of the experimental two microphone transfer function method to compute the radiation impedance of ducts from numerical simulations," 19th International Congress on Sound and Vibration (ICSV19), July 8–12, Vilnius, Lithuania.

Grants and awards

- EAA Best Paper and Presentation Award for young researchers granted by the European Acoustics Association (EAA) during the Forum Acusticum 2014 held in Krakow, Poland.
- EAA grant to attend the Forum Acusticum 2014 held in Krakow, Poland, awarded by the European Acoustics Association (EAA). The grant included free registration to the conference.
- BE-DGR-2013 grant. Predoctoral mobility grant awarded by the government of Catalonia for 6 months of research. This grant was used to make a research stay (18th November 2013 - 5th June 2014) at the Department of Speech, Music and Hearing (TMH) at the Royal Institute of Technology (KTH), Stockholm, Sweden. The local supervisors were Prof. Olov Engwall and Prof. Sten Ternström.
- FI-DGR-2012 grant. Predoctoral grant awarded by the government of Catalonia to perform three years of research (March 2012 - February 2015).

Projects

- EUNISON (Extensive UNIfied-domain SimulatiON of the human voice). This is a European research project of the type FET (Future Emerging Technologies) corresponding to the 7th framework programme that deals with the numerical simulation of the human voice. The project started on the 1st of March 2013 and has a duration of three years. Web: `http://www.eunison.eu/`.

Appendices

Appendix A

Theoretical models for the radiation impedance

In this appendix we include the expressions of some of the models that are commonly used to approximate the radiation impedance of the human mouth and that have been referred thought this work.

A.1 Radiation impedance of a piston set in a baffle

Probably, the most extended theoretical model is that of the impedance of a piston set in an infinite baffle. The resistance R_r and reactance X_r are given by (Kinsler *et al.*, 2000)

$$R_r = 1 - \frac{2J_1(2ka_p)}{2ka_p}, \tag{A.1a}$$

$$X_r = \frac{2\mathrm{H}_1(2ka_p)}{ka_p}, \tag{A.1b}$$

where a_p stands for the radius of the piston, H_1 for the first order Struve function and J_1 for the first order Bessel function of the first kind.

The above expressions are often simplified and only the first terms of a Taylor expansion for $ka << 1$ are taken into account. This is known as the Flanagan approximation (Flanagan, 1972)

$$R_r \simeq \frac{(ka_p)^2}{2}, \tag{A.2a}$$

$$X_r \simeq \frac{8ka_p}{3\pi}. \tag{A.2b}$$

A.2 Radiation impedance of a piston set in a sphere

Another model is that of the piston set in a sphere, which has resistance R_r and reactance X_r values (Morse and Ingard, 1968)

$$R_r = \frac{1}{4}\sum_{m=0}^{\infty} \frac{\left[P_{m-1}(\cos T) - P_{m+1}(\cos T)\right]^2}{(ka_s)^2(2m+1)B_m^2 \sin^2(T/2)}, \tag{A.3a}$$

$$X_r = \frac{-1}{4}\sum_{m=0}^{\infty} \frac{\left[P_{m-1}(\cos T) - P_{m+1}(\cos T)\right]^2}{(2m+1)B_m \sin^2(T/2)} \times \left[j_m(ka_s)\sin(\delta_m) - y_m(ka_s)\cos(\delta_m)\right]. \tag{A.3b}$$

P_m is the Legendre function of order m, j_m and y_m respectively represent the spherical Bessel functions of the first and second kind, a_s is the radius of the sphere and $a_p = 2a_s \sin(T/2)$ is the effective radius of the piston. The radiation impedance magnitude B_m and the phase angle δ_m are given by

$$B_m = \frac{1}{2m+1}\left\{\left[my_{m-1}(ka_s) - (m+1)y_{m+1}(ka_s)\right]^2 + \left[(m+1)j_{m+1}(ka_s) - mj_{m-1}(ka_s)\right]^2\right\}^{1/2}, \tag{A.4a}$$

$$\delta_m = \tan^{-1}\left[\frac{(m+1)j_{m+1}(ka_s) - mj_{m-1}(ka_s)}{my_{m-1}(ka_s) - (m+1)y_{m+1}(ka_s)}\right]. \tag{A.4b}$$

The normalized specific radiation impedance is given by

$$Z_r = R_r + jX_r, \tag{A.5}$$

and the specific acoustic impedance Z' can be obtained as $Z' = Z_0 Z$, with $Z_0 = \rho c \pi a_p^2$ standing for the characteristic impedance of the medium.

A.3 Radiation impedance of a piston set in a circle

In Chapter 7 an analytical expression for the radiation impedance of a two-dimensional vocal tract set in a circle (termed as the piston set in a circle) is needed in an optimization process to attain the acoustic radiation behavior of a three-dimensional vocal tract with circular cross-sections set in a sphere (piston set in a sphere, see Section A.2). This radiation impedance corresponds to a particular case of a piston set in a cylinder (see Sherman, 1958). Considering only a strip of the cylinder and defining $Z_0 = \rho c 2a_p$, it follows that the normalized specific radiation impedance can be computed as

$$Z_r = \frac{2a_c}{\pi a_p} \sum_{m=0}^{\infty} \frac{\sin^2(m\phi)}{m^2} \frac{H_m^{(1)}(ka_c)}{E_m e^{j\gamma_m}}, \tag{A.6}$$

where

$$E_0 = 2\sqrt{J_1^2(ka_c) + Y_1^2(ka_c)}, \qquad m = 0 \tag{A.7a}$$

$$E_m = \frac{1}{2}\left\{[J_{m+1}(ka_c) - J_{m-1}(ka_c)]^2 + [Y_{m-1}(ka_c) - Y_{m+1}(ka_c)]^2\right\}^{1/2}, \qquad m > 0 \tag{A.7b}$$

$$\gamma_0 = \tan^{-1}\left[\frac{-J_1(ka_c)}{Y_1(ka_c)}\right], \qquad m = 0 \tag{A.7c}$$

$$\gamma_m = \tan^{-1}\left[\frac{J_{m+1}(ka_c) - J_{m-1}(ka_c)}{Y_{m-1}(ka_c) - Y_{m+1}(ka_c)}\right], \qquad m > 0 \tag{A.7d}$$

with J_m and Y_m respectively denoting the Bessel functions of the first and second kind.

Appendix B

Waveform glottal models

In this appendix some simple waveform models for generating glottal pulses are described.

B.1 The Rosenberg model

The Rosenberg model of the C type parametrizes the waveform of the volumetric glottal velocity $Q_g(t)$ as

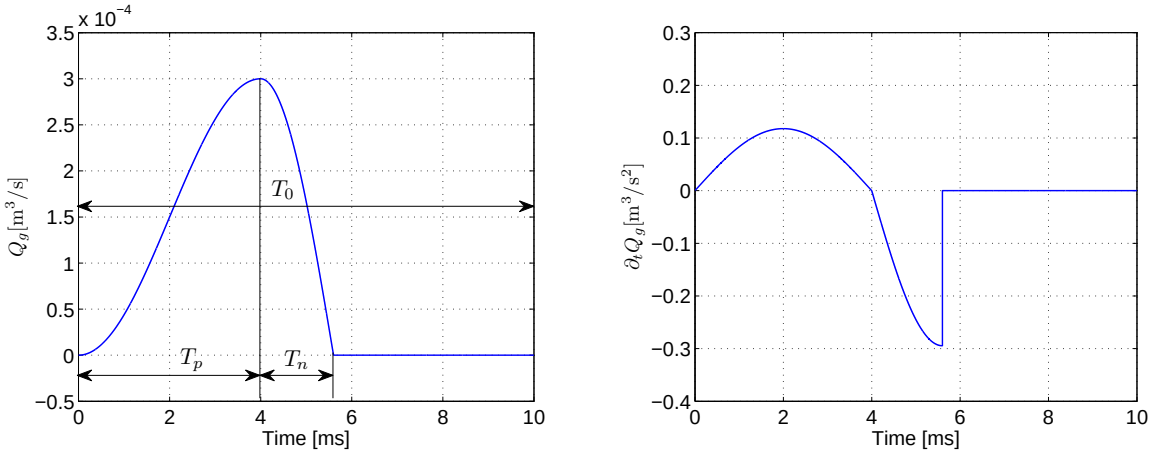
$$Q_g(t) = \begin{cases} \frac{a}{2}\left[1 - \cos\left(\pi \frac{t}{T_p}\right)\right] & 0 \leq t \leq t_p \\ a \cos\left(\frac{\pi}{2} \frac{t - T_p}{T_n}\right) & t_p < t \leq t_p + t_n \\ 0 & t_p + t_n < t < T_0 \end{cases} , \tag{B.1}$$

where a is the amplitude, t_p is the positive slope, t_n the negative slope and T_0 the period of the glottal pulses (see Fig. B.1). The times t_p and t_n are related to the fundamental frequency $F_0 = 1/T_0$ as

$$t_p = 40\% \; T_0, \tag{B.2}$$

$$t_n = 16\% \; T_0. \tag{B.3}$$

In Fig. B.1 one pulse generated by the C Rosenberg model is shown, using $a = 3 \times 10^{-4} \mathrm{m}^3/\mathrm{s}^2$ and $F_0 = 100$ Hz. However, this pulse contains high frequency information because it has abrupt transitions (i.e. it is not smooth). In numerical simulations, high frequency information can pollute the solution because these can not be captured by the mesh (the size of the elements can not be fine enough to capture it). This may result in errors and instabilities of the solution.



(a) (b)

Fig. B.1: (a) Volume velocity $Q_g(t)$ and (b) its time derivative $\partial_t Q_g$ generated by the Rosenberg model of the C type. In (a) the times T_0, T_p and T_n that characterize the pulse $Q_g(t)$ have been indicated.

B.2 The LF model

The LF model (Fant *et al.*, 1985) also describes the flow variations of the glottal source, but it does so acting on its time derivative. This is an important feature because imposing the time derivative of the glottal flow is a natural way to couple the source model with the vocal tract acoustic model, see Eq. (2.3). However, the most important advantage is that this model describes a smooth pulse, closer to reality than the Rosenberg one. Therefore, in contrast to the Rosenberg pulses, no excessive high frequency is expected to be introduced into the numerical simulations.

In the LF model, the time derivative of the volume velocity $Q_g(t)$ is given by

$$\begin{cases} E_0 e^{\alpha t} \sin(w_g t) & 0 \le t \le t_e, \\ -\frac{E_e}{\epsilon t_a}\left(e^{-\epsilon(t-t_e)} - e^{-\epsilon(T_0-te)}\right) & t_e < T_0, \end{cases} \tag{B.4}$$

where T_0 is the inverse of the fundamental frequency F_0. Prior to defining parameters E_0, α, w_g and t_e, it is worth introducing the following dimensionless quantities

$$R_a = \frac{t_a}{T_0 - t_e}, \quad R_k = \frac{t_e - t_p}{t_p}, \quad R_g = \frac{T_0}{2t_p}. \tag{B.5}$$

The definitions of t_e, t_p, t_a, T_0 and E_e are given in Fig. B.2. The angular frequency w_g is related to the fundamental frequency F_0 by $w_g = 2\pi F_0 R_g$. For a male voice,

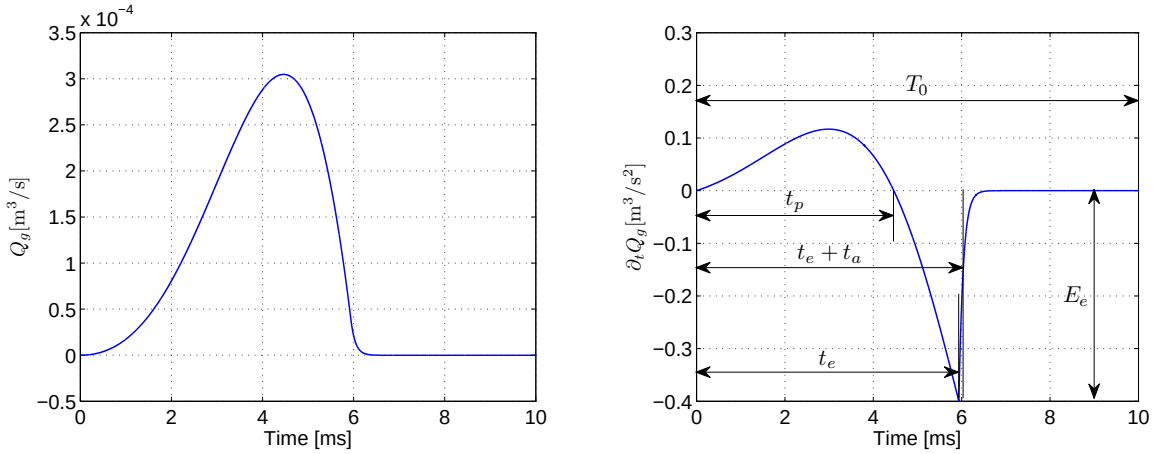


Fig. B.2: (a) Volume velocity $Q_g(t)$ and (b) its time derivative $\partial_t Q_g(t)$ generated by the LF model. In (b) the four wave shape parameters t_p, t_e, t_a and E_e that characterize the pulse $\partial_t Q_g(t)$ have been indicated.

$E_e = 0.4$ m^{3}s^{-2}, $R_g = 1.12$ and $R_k = 0.34$ (see e.g., Vampola *et al.*, 2008b). The time t_a, for small values of ϵ can be computed as

$$t_a = \frac{U_e}{E_e}, \tag{B.6}$$

where U_e is the mean flow volume rate in the glottis that for a male has a value of $U_e = 0.12$ l/s (Vampola *et al.*, 2008b). Then, using (B.6) we can compute the time $t_a = 0.3$ ms. On the other hand, if we impose that the time derivative of the glottal pulse (B.4) at time t_p must be zero and that it has to be continuous at time t_e, we are left with the following non-linear equations

$$E_0 e^{\alpha t_p} \sin(w_g t_p) = 0, \tag{B.7a}$$

$$E_0 e^{\alpha t_e} \sin(w_g t_e) = E_e, \tag{B.7b}$$

$$-\frac{E_e}{\epsilon t_a}\left(1 - e^{-\epsilon(T_0 - te)}\right) = E_e. \tag{B.7c}$$

The solution to this system provides the parameters E_e, α and ϵ.

Bibliography

Aalto, D., Aaltonen, O., Happonen, R.-P., Jääsaari, P., Kivelä, A., Kuortti, J., Luukinen, J.-M., Malinen, J., Murtola, T., Parkkola, R., Saunavaara, J., T. Soukka, T., and Vainio, M. (**2014**), "Large scale data acquisition of simultaneous MRI and speech," Appl. Acoust., **83**, pp. 64–75.

Abramowitz, M. and Stegun, I. (**1970**), *Handbook of Mathematical Functions* (Dover, New York).

Al Bawab, Z., Raj, B., and Stern, R. M. (**2010**), "A hybrid physical and statistical dynamic articulatory framework incorporating analysis-by-synthesis for improved phone classification," in *Proc. of ICASSP'10* (Dallas, Texas, USA), pp. 4194–4197.

Apostol, L., Perrier, P., and Bailly, G. (**2004**), "A model of acoustic interspeaker variability based on the concept of formant-cavity affiliation," J. Acoust. Soc. Am., **115**(1), pp. 337–351.

Arnela, M. and Guasch, O. (**2012**), "Adaptation of the experimental two-microphone transfer function method to compute the radiation impedance of ducts from numerical simulations," in *Proc. of ICSV19* (Vilnius, Lithuania).

Arnela, M. and Guasch, O. (**2013**), "Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method," J. Acoust. Soc. Am., **133**(6), pp. 4197–4209.

Arnela, M. and Guasch, O. (**2014**a), "Three-dimensional behavior in the numerical generation of vowels using tuned two-dimensional vocal tracts," in *Proc. of 7th Forum Acousticum* (Kraków, Poland).

Arnela, M. and Guasch, O. (**2014**b), "Two-dimensional vocal tracts with three-dimensional behaviour in the numerical production of vowels," J. Acoust. Soc. Am., **135**(1), pp. 369–379.

Arnela, M. and Guasch, O. (**2014**c), "Validation of the piston set in a sphere model for vowel sound radiation losses against realistic head geometry using time-domain finite-element simulations," in *9th International Conference on Voice Physiology and Biomechanics (ICVPB)* (Salt Lake City, USA).

Arnela, M., Guasch, O., and Alías, F. (**2012**), "Analysis of the radiation effects on vowels by means of time domain finite element simulations," in *Proc. of ICSV19* (Vilnius, Lithuania).

Arnela, M., Guasch, O., and Alías, F. (**2013**), "Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations," J. Acoust. Soc. Am., **134**(4), pp. 2946–2954.

Arnela, M., Guasch, O., Codina, R., and Espinoza, H. (**2014**), "Finite element computation of diphthong sounds using tuned two-dimensional vocal tracts," in *Proc. of 7th Forum Acousticum* (Kraków, Poland).

Badia, S., Codina, R., and Espinoza, H. (**2014**), "Stability, convergence and accuracy of stabilized finite elements methods for the wave equation in mixed form," SIAM J. Numer. Anal., **52**(4), pp. 1729–1752.

Bazilevs, Y., Calo, V. M., Cottrell, J. A., Hughes, T. J. R., Reali, A., and Scovazzi, G. (**2007**), "Variational multiscale residual-based turbulence modeling for large eddy simulation," Comput. Methods Appl. Mech. Engrg., **197**(1–4), pp. 173–201.

Berenger, J. P. (**1994**), "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys., **114**(2), pp. 185–200.

Birkholz, P. (**2013**), "Modeling Consonant-Vowel Coarticulation for Articulatory Speech Synthesis," PLoS ONE, **8**(4), pp. e60603. doi:10.1371/journal.pone.0060603.

Birkholz, P. and Jackèl, D. (**2003**), "A three-dimensional model of the vocal tract for speech synthesis," in *15th ICPhS* (Barcelona, Spain), pp. 2597–2600.

Birkholz, P., Jackèl, D., and Kröger, B. J. (**2006**), "Construction and control of a three-dimensional vocal tract model," in *Proc. of ICASSP'06* (Toulouse, France), pp. 873–876.

Birkholz, P., Jackèl, D., and Kröger, B. J. (**2007**), "Simulation of losses due to turbulence in the time-varying vocal system," IEEE Trans. Audio Speech Lang. Process., **15**(4), pp. 1218–1225.

Birkholz, P. and Kröger, B. J. (**2006**), "Vocal tract model adaptation using magnetic resonance imaging," in *Proc. of ISSP'06* (Ubatuba, Brazil), pp. 493–500.

Bitmapworld (**2005**), "Head + Morph Targets 3D model," (Turbosquid, New Orleans, LA), availabe online at http://www.turbosquid.com/3d-models/3d-model-male-head-morph-targets/261694 (Last viewed 24 October 2014), .

Blandin, R., Arnela, M., Laboissière, R., Pelorson, X., Van Hirtum, A., and Guasch, O. (**2014**a), "Effects of higher order propagation modes in vocal tract like geometries," J. Acoust. Soc. Am., Accepted.

Blandin, R., Pelorson, X., Van Hirtum, A., Laboissière, R., Arnela, M., and Guasch, O. (**2014**b), "Modélisation aéroacoustique de la production de la parole," in *Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal audio* (Lyon, France).

Blandin, R., Pelorson, X., Van Hirtum, A., Laboissière, R., Guasch, O., and Arnela, M. (**2014**c), "Effet des modes de propagation non plan dans les guides d'ondes à section variable," in *French Acoustical Conference 2014* (Poitiers, France), pp. 745–751.

Boij, S. and Nilsson, B. (**2003**), "Reflection of sound at area expansions in a flow duct," J. Sound Vib., **260**(3), pp. 477 – 498.

Boonen, R., Sas, P., Desmet, W., Lauriks, W., and Vermeir, G. (**2009**), "Calibration of the two microphone transfer function method with hard wall impedance measurements at different reference sections," Mech. Syst. Signal Pr., **23**(5), pp. 1662–1671.

Burkhard, M. D. and Sachs, R. M. (**1975**), "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am., **58**(1), pp. 214–222.

Chalker, D. A. and Mackerras, D. (**1985**), "Models for representing the acoustic radiation impedance of the mouth," IEEE Trans. Acoustics Speech Signal Process., **33**(6), pp. 1606–1609.

Chung, J. Y. and Blaser, D. A. (**1980**), "Transfer-function method of measuring in-duct acoustic properties. I. Theory," J. Acoust. Soc. Am., **68**(3), pp. 907–913.

Codina, R. (**1998**), "Comparison of some finite element methods for solving the diffusion-convection-reaction equation," Comput. Methods Appl. Mech. Engrg., **156**(1–4), pp. 185–210.

Codina, R. (**2002**), "Stabilized finite element approximation of transient incompressible flows using orthogonal subscales," Comput. Methods Appl. Mech. Engrg., **191**(39–40), pp. 4295–4321.

Codina, R. (**2008**), "Finite element approximation of the hyperbolic wave equation in mixed form," Comput. Methods Appl. Mech. Engrg., **197**(13–16), pp. 1305–1322.

Codina, R., González-Ondina, J., Díaz-Hernández, G., and Principe, J. (**2008**), "Finite element approximation of the modified Boussinesq equations using a stabilized formulation," Int. J. Numer. Meth. Fluids, **57**(9), pp. 1249–1268.

Codina, R., Principe, J., Guasch, O., and Badia, S. (**2007**), "Time dependent subscsales in the stabilized finite element approximation of incompressible flow problems," Comput. Methods Appl. Mech. Engrg., **196**(21–24), pp. 2413–2430.

Dalmont, J.-P., Nederveen, C. J., and Joly, N. (**2001**), "Radiation impedance of tubes with different flanges: numerical and experimental investigations," J. Sound Vib., **244**(3), pp. 505–534.

Dang, J. and Honda, K. (**2004**), "Construction and control of a physiological articulatory model," J. Acoust. Soc. Am., **115**(2), pp. 853–870.

Doel, K. v. d. and Ascher, U. (**2008**), "Real-Time Numerical Solution of Webster's Equation on A Nonuniform Grid," IEEE Trans. Audio Speech Lang. Process., **16**(6), pp. 1163–1172.

Donea, J., Huerta, A., Ponthot, J.-P., and Rodríguez-Ferran, A. (**2004**), *Arbitrary Lagrangian Eulerian Methods, in: Encyclopedia of Computational Mechanics* (John Wiley & Sons, Ltd).

Escolano, J., Spa, C., Garriga, A., and Mateos, T. (**2013**), "Removal of afterglow effects in 2-D discrete-time room acoustics simulations," Appl. Acoust., **74**(6), pp. 818–822.

Espinoza, H., Codina, R., and Badia, S. (**2014**), "A Sommerfeld non-reflecting boundary condition for the wave equation in mixed form," Comput. Methods Appl. Mech. Engrg., **276**(0), pp. 122–148.

Fabre, B., Gilbert, J., Hirschberg, A., and Pelorson, X. (**2012**), "Aeroacoustics of Musical Instruments," Annual Review of Fluid Mechanics, **44**(1), pp. 1–25.

Fahy, F. (**2001**), *Foundations of Engineering Acoustics* (Academic Press, New York).

Fant, G. (**1970**), *Acoustic Theory of Speech Production* (Mouton, Paris), 2nd ed.

Fant, G., Liljencrants, J., and Lin, Q. G. (**1985**), "A four-parameter model of glottal flow," STL-QPSR, **77**(4), pp. 1–13.

Fels, S., Stavness, I., Hannam, A. G., Lloyd, J. E., Anderson, P., Batty, C., Chen, H., Combe, C., Pang, T., Mandal, T., Teixeira, B., Green, S., Bridson, R., Lowe, A., Almeida, F., Fleetham, J., and Abugharbieh, R. (**2009**), "Advanced Tools for Biomechanical Modeling of the Oral, Pharyngeal, and Laryngeal Complex,"

in *International Symposium on Biomechanics Healthcare and Information Science, electronic proceedings*, pp. 2725–2728.

Flanagan, J. L. (**1972**), *Speech Analysis, Synthesis and Perception* (Springer-Verlag, New York), 2nd ed.

Flanagan, J. L. (**2008**), *Speech Analysis, Synthesis and Perception* (Springer-Verlag, New York), 3th ed.

Flanagan, J. L. and Landgraf, L. (**1968**), "Self-Oscillating Source for Vocal-Tract Synthesizers," IEEE Trans. Audio Electroacustics, **16**(1), pp. 57–64.

Fleischer, M., Pinkert, S., Mattheus, W., Mainka, A., and Mürbe, D. (**2014**), "Formant frequencies and bandwidths of the vocal tract transfer function are affected by the mechanical impedance of the vocal tract wall," Biomech. Model. Mechanobiol., pp. 1–15.

Fletcher, N. H. and Rossing, T. D. (**1988**), *The Physics of Musical Instruments* (Springer-Verlag, New York).

Fromkin, V. (**1964**), "Lip positions in american english vowels," Language and speech, **7**(4), pp. 215–225.

Grote, M. J. and Sim, I. (**2010**), "Efficient PML for the wave equation," Global Science Preprint, arXiv:1001.0319v1 [math.NA], pp. 1–15.

Guasch, O. (**2009**), "Direct transfer functions and path blocking in a discrete mechanical system," J. Sound Vib., **321**(3-5), pp. 854–874.

Guasch, O., Arnela, M., Codina, R., and Espinoza, H. (**2014**a), "A stabilized finite element method for the mixed wave equation in an ALE framework with application to diphthong production," J. Comput. Phys., Submitted.

Guasch, O. and Codina, R. (**2007**), "An algebraic subgrid scale finite element method for the convected Helmholtz equation in two dimensions with applications in aeroacoustics," Comput. Methods Appl. Mech. Engrg., **196**(45–48), pp. 4672–4689.

Guasch, O., Codina, R., Arnela, M., and Espinoza, H. (**2014**b), "A stabilized Arbitrary Lagrangian Eulerian Finite Element Method for the mixed wave equation with application to diphthong production," in *11th World Congress on Computational Mechanics (WCCM)* (Barcelona, Spain).

Guasch, O. and Magrans, F. X. (**2004**), "The Global Transfer Direct Transfer method applied to a finite simply supported elastic beam," J. Sound Vib., **276**(1–2), pp. 335–359.

Guasch, O., Ternström, S., Arnela, M., and Alías, F. (**2013**), "Unified numerical simulation of the physics of voice. The EUNISON project." in *Proc. of 8th ISCA Speech Synthesis Workshop* (Aug 31-Sept 2, 2013, Barcelona, Catalonia, Spain), pp. 241–242.

Gutiérrez-Vega, J. C. (**2000**), "Formal analysis of the propagation of invariant optical fields in elliptic coordinates," Ph.D. dissertation, Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México.

Hannukainen, A., Lukkari, T., Malinen, J., and Palo, P. (**2007**), "Vowel formants from the wave equation," J. Acoust. Soc. Am., **122**(1), pp. EL1–EL7.

Hauke, G. (**1998**), "A symmetric formulation for computing transient shallow water flows," Comput. Methods Appl. Mech. Engrg., **163**(1–4), pp. 111–122.

Hirt, C. W., Amsden, A. A., and Cook, J. L. (**1974**), "An arbitrary Lagrangian-Eulerian computing method for all flow speeds," J. Comput. Phys., **14**(3), pp. 227–254.

Howard, D. M., Ternström, S., and Speed, M. (**2009**), "Natural Voice Synthesis: The potential relevance of high-frequency components," in *Proc. of 3rd Advanced Voice Function Assessment International Workshop* (Madrid, Spain).

Howe, M. S. and McGowan, R. S. (**2007**), "Sound generated by aerodynamic sources near a deformable body, with application to voiced speech," J. Fluid Mech., **592**, pp. 367–392.

Huang, A. and Temam, R. (**2014**), "The linearized 2D inviscid shallow water equations in a rectangle: boundary conditions and well-posedness," Archives for Rational Mechanics and Analysis, **211**(3), pp. 1027–1063.

Huerta, A. and Liu, W. K. (**1988**), "Viscous flow with large free surface motion," Comput. Methods Appl. Mech. Engrg., **69**(3), pp. 277–324.

Hughes, T. J. R. (**1995**), "Multiscale phenomena: Green's function, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized formulations," Comput. Methods Appl. Mech. Engrg., **127**(1), pp. 387–401.

Hughes, T. J. R. (**2000**), *The finite element method. Linear Static and Dynamic Finite Element Analysis* (Dover, New York).

Hughes, T. J. R., Feijóo, G. R., Mazzei, L., and Quincy, J. B. (**1998**), "The variational multiscale method, a paradigm for computational mechanics," Comput. Methods Appl. Mech. Engrg., **166**(1), pp. 3–24.

Hughes, T. J. R., Liu, W. K., and Zimmermann, T. K. (**1981**), "Lagrangian-Eulerian finite-element formulation for compressible viscous flows," Comput. Methods Appl. Mech. Engrg., **29**(3), pp. 329–349.

Ihlenburg, F. (**1998**), *Finite Element Analysis of Acoustic Scattering, Applied Mathematical Sciences* (Springer, Berlin).

Ishizaka, K. and Flanagan, J. L. (**1972**), "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. Journal, **51**(6), pp. 1233–1268.

Ishizaka, K., French, J., and Flanagan, J. L. (**1975**), "Direct determination of vocal tract wall impedance," IEEE Trans. Acoustics Speech Signal Process., **23**(4), pp. 370–373.

ISO 10534-2 (**1998**), *Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes—Part 2: Transfer-function method* (International Organisation for Standardization, Geneva, Switzerland).

Jang, S.-H. and Ih, J.-G. (**1998**), "On the multiple microphone method for measuring in-duct acoustic properties in the presence of mean flow," J. Acoust. Soc. Am., **103**(3), pp. 1520–1526.

Kagawa, Y., Shimoyama, R., Yamabuchi, T., Murai, T., and Takarada, K. (**1992**), "Boundary element models of the vocal tract and radiation field and their response characteristics," J. Sound Vib., **157**(3), pp. 385–403.

Kako, T. and Touda, K. (**2006**), "Numerical method for voice generation problem based on finite element method," J. Comput. Acoust., **14**(1), pp. 45–56.

Kaltenbacher, B., Kaltenbacher, M., and Sim, I. (**2013**), "A modified and stable version of a perfectly matched layer technique for the 3-d second order wave equation in time domain with an application to aeroacoustics," J. Comput. Phys., **235**, pp. 407–422.

Kaltenbacher, M., Zörner, S., Hüppe, A., and Sidlof, P. (**2014**), "3D numerical simulations of human phonation," in *11th World Congress on Computational Mechanics (WCCM)* (Barcelona, Spain), pp. 1–11.

Kang, Z. and Ji, Z. (**2008**), "Acoustic length correction of duct extension into a cylindrical chamber," J. Sound Vib., **310**(4–5), pp. 782–791.

Kelly, J. and Lochbaum, C. (**1962**), "Speech synthesis," in *Proc. Fourth ICA* (Copenhagen, Denmark), pp. 1–4.

Kemp, J. A., Campbell, D. M., and Amir, N. (**2001**), "Multimodal radiation impedance of a rectangular duct terminated in an infinite baffle," Acta Acust. united Ac., **87**(1), pp. 11–15.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**), *Fundamentals of Acoustics* (Wiley, New York), 4th ed.

Kröger, B. J. and Birkholz, P. (**2009**), *Articulatory Synthesis of Speech and Singing: State of the Art and Suggestions for Future Research, A. Esposito et al. (Eds.): Multimodal Signals: Cognitive and Algorithmic Issues* (Springer), vol. 5398/2009, pp. 306–319.

Langtangen, H. P. and Pedersen, G. (**1998**), "Computational models for weakly dispersive nonlinear water waves," Comput. Methods Appl. Mech. Engrg., **160**(3–4), pp. 337–358.

Link, G., Kaltenbacher, M., Breuer, M., and Döllinger, M. (**2009**), "A 2D Finite-element scheme for fluid-solid-acoustic interactions and its application to human phonation," Comput. Methods Appl. Mech. Engrg., **198**(41–44), pp. 3321–3334.

Lous, N. J. C., Hofmans, G. C. J., Veldhuis, R. N. J., and Hirschberg, A. (**1998**), "A Symmetrical Two-Mass Vocal-Fold Model Coupled to Vocal Tract and Trachea, with Application to Prosthesis Design," Acta Acust. united Ac., **84**(6), pp. 1135–1150.

Lowson, M. V. and Baskaran, S. (**1975**), "Propagation of sound in elliptic ducts," J. Sound Vib., **38**(2), pp. 185–194.

Matsuzaki, H., Miki, N., and Ogawa, Y. (**2000**), "3D finite element analysis of Japanese vowels in elliptic sound tube model," Electron. Comm. Jpn. 3, **83**(4), pp. 43–51.

Matsuzaki, H. and Motoki, K. (**2007**), "Study of acoustic characteristics of vocal tract with nasal cavity during phonation of Japanese /a/," Acoust. Sci. & Tech., **28**(2), pp. 124–127.

Matsuzaki, H., Serrurier, A., Badin, P., and Motoki, K. (**2008**), "Time-domain FEM simulation of Japanese and French vowel /a/ with nasal coupling," in *Proc. of Spring Meeting of the Acoustical Society of Japan* (Japan), no. 2–4–9, pp. 331–332.

McGowan, R. S. and Howe, M. S. (**2007**), "Compact Green's functions extend the acoustic theory of speech production," J. Phonetics, **35**(2), pp. 259–270.

Mermelstein, P. (**1973**), "Articulatory model for the study of speech production," J. Acoust. Soc. Am., **53**(4), pp. 1070–1082.

Mokhtari, P., Takemoto, H., Nishimura, R., and Kato, H. (**2008**), "Computer Simulation of HRTFs for Personalization of 3D Audio," in *Proc. of ISUC'08* (Osaka, Japan), pp. 435–440.

Monson, B. B., Lotto, A. J., and Ternström, S. (**2011**), "Detection of high-frequency energy changes in sustained vowels produced by singers," J. Acoust. Soc. Am., **129**(4), pp. 2263–2268.

Morfey, C. L. (**1969**), "A note on the radiation efficiency of acoustic duct modes," J. Sound Vib., **9**(3), pp. 367–372.

Morse, P. M. and Ingard, K. U. (**1968**), *Theoretical Acoustics* (McGraw-Hill, New York).

Motoki, K. (**2002**), "Three-dimensional acoustic field in vocal-tract," Acoust. Sci. & Tech., **23**(4), pp. 207–212.

Motoki, K., Badin, P., Pelorson, X., and Matsuzaki, H. (**2000**), "A modal parametric method for computing acoustic characteristics of three-dimensional vocal tract models," in *Proc. of 5th Seminar on Speech Production: Models and Data*, pp. 325–328.

Mullen, J., Howard, D. M., and Murphy, D. T. (**2006**), "Waveguide physical modeling of vocal tract acoustics: flexible formant bandwidth control from increased model dimensionality," IEEE Trans. Audio Speech Lang. Process., **14**(3), pp. 964–971.

Mullen, J., Howard, D. M., and Murphy, D. T. (**2007**), "Real-Time Dynamic Articulations in the 2-D Waveguide Mesh Vocal Tract Model," IEEE Trans. Audio Speech Lang. Process., **15**(2), pp. 577–585.

Munjal, M. L. (**1987**), *Acoustics of Ducts and Mufflers* (Wiley, New York).

Nieuwenhof, B. v. d. and Coyette, J.-P. (**2001**), "Treatment of frequency-dependent admittance boundary conditions in transient acoustic finite/infinite-element models," J. Acoust. Soc. Am., **110**(4), pp. 1743–1751.

Oliveira, J. M. G. S. and Gil, P. J. S. (**2010**), "Propagation of sound in ducts with elliptical cross-section and lined walls," in *Proc. of Inter-noise'10* (Lisbon, Portugal).

Palo, J., Aalto, D., Aaltonen, O., Happonen, R.-P., Malinen, J., Saunavaara, J., and Vainio, M. (**2012**), "Articulating Finnish Vowels: Results from MRI and sound data," Linguistica Uralica, **48**(3), pp. 194–199.

Pelorson, X., Hirschberg, A., Van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**), "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am., **96**(6), pp. 3416–3431.

Pelorson, X., Van Hirtum, A., Mondet, B., Guasch, O., and Arnela, M. (**2013**), "Three-dimensional vocal tract acoustics," in *Acoustics 2013* (New Delhi, India).

Ridouane, R. (**2006**), "Investigating speech production: A review of some techniques," Laboratoire de Phonétique et Phonologie, [Online Document] Available at http://lpp.univ-paris3.fr/equipe/rachid_ridouane/Ridouane_Investigating.pdf.

Rosenberg, A. E. (**1971**), "Effect of Glottal Pulse Shape on the Quality of Natural Vowels," J. Acoust. Soc. Am., **49**(2), pp. 583–590.

Selamet, A., Ji, Z. L., and Kach, R. A. (**2001**), "Wave reflections from duct terminations," J. Acoust. Soc. Am., **109**(4), pp. 1304–1311.

Serrurier, A. and Badin, P. (**2008**), "A three-dimensional articulatory model of the velum and nasopharyngeal wall based on MRI and CT data," J. Acoust. Soc. Am., **123**(4), pp. 2335–2355.

Sherman, C. H. (**1958**), "Mutual Radiation Impedance Between Pistons on Spheres and Cylinders," Tech. Rep. No. 405, U.S. Navy underwater sound laboratory, Fort Trumbull, New london, Conn.

Sivian, L. J. (**1937**), "Sound Propagation in Ducts Lined with Absorbing Materials," J. Acoust. Soc. Am., **9**(1), pp. 77–77.

Sondhi, M. M. (**1983**), "An improved vocal tract model," in *Proc. of the 11th ICA* (Paris, France), pp. 167–170.

Sondhi, M. M. and Schroeter, J. (**1987**), "A hybrid time-frequency domain articulatory speech synthesizer," IEEE Trans. Audio Speech Lang. Process., **35**(7), pp. 955–967.

Spa, C., Escolano, J., Garriga, A., and Mateos, T. (**2010**), "Compensation of the Afterglow Phenomenon in 2-D Discrete-Time Simulations," IEEE Signal Process. Letters, **17**(8), pp. 758–761.

Speed, M., Murphy, D. T., and Howard, D. M. (**2009**), "Characteristics of two-dimensional finite difference techniques for vocal tract analysis and voice synthesis," in *Proc. of Interspeech'09* (Brighton, United Kingdom), pp. 768–771.

Speed, M., Murphy, D. T., and Howard, D. M. (**2013**), "Three-Dimensional Digital Waveguide Mesh Simulation of Cylindrical Vocal Tract Analogs," IEEE Trans. Audio Speech Lang. Process., **21**(2), pp. 449–455.

Stark, J., Ericsdotter, C., Branderud, P., Branderurd, P., Sundberg, J., and H. Lundberg, J. L. (**1999**), "The Apex Model as a tool in the specification of speakerspecific articulatory behavior," in *Proc. from the XIVth ICPhS* (San Francisco, USA), pp. 1–7.

Stavness, I., Lloyd, J. E., Payan, Y., and Fels, S. (**2011**), "Coupled hard-soft tissue simulation with contact and constraints applied to jaw-tongue-hyoid dynamics," Int. J. Numer. Method. Biomed. Eng. **27**(3), pp. 367–390.

Stevens, K. N. (**2000**), *Acoustic phonetics* (The MIT Press, Cambridge, MA).

Story, B. H. (**2002**), "An overview of the physiology, physics and modeling of the sound source for vowels," Acoustical Science and Technology, **23**(4), pp. 195–206.

Story, B. H. (**2005**), "A parametric model of the vocal tract area function for vowel and consonant simulation," J. Acoust. Soc. Am., **117**(5), pp. 3231–3254.

Story, B. H. (**2006**), "Technique for "tuning" vocal tract area functions based on acoustic sensitivity functions," J. Acoust. Soc. Am., **119**(2), pp. 715–718.

Story, B. H. (**2008**), "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002," J. Acoust. Soc. Am., **123**(1), pp. 327–335.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**), "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am., **100**(1), pp. 537–554.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**2001**), "The relationship of vocal tract shape to three voice qualities," J. Acoust. Soc. Am., **109**(4), pp. 1651–1667.

Švancara, P. and Horáček, J. (**2006**), "Numerical modelling of effect of tonsillectomy on production of czech vowels," Acta Acust. united Ac., **92**(5), pp. 681–688.

Švancara, P., Horáček, J., Vokrál, J., and Cerný, L. (**2006**), "Computational modelling of effect of tonsillectomy on voice production," Logoped. Phoniatr. Vocol., **31**(3), pp. 117–125.

Takemoto, H., Adachi, S., Mokhtari, P., and Kitamura, T. (**2013**), "Acoustic interaction between the right and left piriform fossae in generating spectral dips," J. Acoust. Soc. Am., **134**(4), pp. 2955–2964.

Takemoto, H., Honda, K., Masaki, S., Shimada, Y., and Fujimoto, I. (**2006**), "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," J. Acoust. Soc. Am., **119**(2), pp. 1037–1049.

Takemoto, H., Mokhtari, P., and Kitamura, T. (**2010**), "Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method," J. Acoust. Soc. Am., **128**(6), pp. 3724–3738.

Tian, F.-B., Dai, H., Luo, H., Doyle, J. F., and Rousseau, B. (**2014**), "Fluid-structure interaction involving large deformations: 3D simulations and applications to biological systems," J. Comput. Phys., **258**(0), pp. 451–469.

Vampola, T., Horáček, J., Laukkanen, A. M., and Švec, J. G. (**2013**), "Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modelling based on CT measurement," Logopedics Phoniatrics Vocology, pp. 1–10.

Vampola, T., Horáček, J., and Švec, J. G. (**2008**a), "FE Modeling of Human Vocal Tract Acoustics. Part I: Production of Czech vowels," Acta Acust. united Ac., **94**(5), pp. 433–447.

Vampola, T., Horáček, J., Vokářál, J., and Černý, L. (**2008**b), "FE Modeling of Human Vocal Tract Acoustics. Part II: Influence of velopharyngeal insufficiency on phonation of vowels," Acta Acust. united Ac., **94**(5), pp. 448–460.

Vampola, T., Laukkanen, A. M., Horáček, J., and Švec, J. G. (**2011**), "Vocal tract changes caused by phonation into a tube: a case study using computer tomography and finite-element modeling," J. Acoust. Soc. Am., **129**(1), pp. 310–315.

Van Hirtum, A. and Fujiso, Y. (**2012**), "Insulation room for aero-acoustic experiments at moderate Reynolds and low Mach numbers," Appl. Acoust., **73**(1), pp. 72–77.

Walkley, M. and Berzins, M. (**2002**), "A finite element method for the two-dimensional extended Boussinesq equations," Int. J. Numer. Meth. Fluids, **39**(10), pp. 865–885.

Wall, W. A., Gerstenberger, A., Gammnitzer, P., Förster, C., and Ramm, E. (**2006**), *Large deformation fluid-structure interaction − advances in ALE methods and new fixed grids approaches, in: H.-J. Bungrartz, M. Schäfer (Eds.), Fluid-structure interaction: modeling, simulation, optimization, LNCSE* (Springer, Heidelberg).

Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (**2002**a), "Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms," J. Acoust. Soc. Am., **112**(5), pp. 2134–2146.

Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (**2002**b), "Computational aeroacoustics of phonation, Part II: Effects of flow parameters and ventricular folds," J. Acoust. Soc. Am., **112**(5), pp. 2147–2154.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., and Choe, A. (**2008**), "A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/," J. Acoust. Soc. Am., **123**(6), pp. 4466–4481.

Zorumski, W. E. (**1973**), "Generalized radiation impedances and reflection coefficients of circular and annular ducts," J. Acoust. Soc. Am., **54**(6), pp. 1667–1673.